VYTAUTO DIDŽIOJO UNIVERSITETAS

INFORMATIKOS FAKULTETAS

MATEMATIKOS IR STATISTIKOS KATEDRA

Mindaugas Indriūnas

BRANDUOLIŲ METODŲ TAIKYMAI KLASIFIKAVIMO UŽDAVINIAMS KORONARŲ STENOZĖS PROGNOZAVIME

Bakalauro baigiamasis darbas

Matematikos ir jos taikymų studijų programa, valstybinis kodas 61201P105

Fizinių mokslų studijų kryptis

Vadovė doc.dr. Jonė Venclovienė , 2010-06-11

(parašas)

Apginta doc.dr. Kęstutis Šidlauskas , 2010-06-11

(parašas)

KAUNAS 2010

Turinys

SANTRAUKA

Bakalauro darbo autorius:	Mindaugas Indriūnas
Bakalauro darbo pavadinimas:	Branduolių teorijos taikymai klasifikavimo uždaviniams koronarų stenozės prognozavime
Vadovas:	Doc. Dr. Jonė Venclovienė
Darbas pristatytas:	Vytauto Didžiojo Universitetas, Informatikos fakultetas, Kaunas, 2010 birželis.
Puslapių skaičius:	38
Lentelių skaičius:	3
Paveikslų skaičius:	5
Priedų skaičius:	0

Šiame darbe formuluojamas klasifikavimo teorijos keliamas uždavinys, statistinio klasifikavimo uždavinys; trumpai susipažįstama su klasifikatoriais, ir gilinamasi į požymių išskyrimą; susipažįstama su pagrindinių komponenčių metodo išvedimu maksimizuojant dispersiją Lagranžo daugiklių metodu; apžvelgiami savireprodukuojančių branduolių teorijos pagrindiniai teiginiai ir jų įrodymai; susipažįstama su Hilberto-Šmidto nepriklausomumo kriterijumi, apžvelgiami du šio kriterijaus įverčiais ($HSIC_0$, $HSIC_1$) paremti požymių išskyrimo metodai ir jų išvedimas. Išnagrinėjamas įverčiu $HSIC_0$ paremtas metodas ir randama jame naudojamos matricos dekompoziciją. Pateikiami šiais metodais gauto klasifikavimo su k-artimiausių kaimynų ir Logistinės regresijos klasifikatoriais rezultatai.

ABSTRACT

Author of Bachelor Thesis:	Mindaugas Indriūnas
Full title of Bachelor Thesis:	Applications of Kernel Methods in Classification Tasks in Coronary Artery Stenosis Prognosis
Supervisor:	Dr., Assoc. Prof. Jonė Vencloviėnė
Presented at:	Vytautas Magnus University, Faculty of Informatics, Kaunas, June 2010.
Number of pages:	38
Number of tables:	3
Number of pictures:	5
Number of appendices:	0

In this work we study the problem of classification. We begin from studying the problem of pattern recognition and statistical classification, study classifiers, and feature extraction; study the derivation of principal component analysis method by maximizing the variance with the method of Lagrange multipliers; overview kernel methods and study the proofs of the reproducing kernel theory; study the Hilbert-Schmidt Independence Criterion and the derivation of component analysis based on two of HSIC estimates (biased and unbiased), and finding the decomposition of the matrix used in the case of biased estimate. Providing the classification results with k-nearest neighbor and Logistic regression classifiers.

1. ĮVADAS

1.1. Atpažinimo teorijos uždavinys

Atpažinimo teorija (*pattern recognition, pattern classification*) nagrinėja duomenų klasifikavimo uždavinį. Pačiu bendriausiu atveju, turime duomenis, ir nieko daugiau apie juos nežinodami ieškome klasių, kurias tie duomenys apibrėžia savo vidine struktūra. Pavyzdžiui, turime taškų aibę $\{\mathbf{x}_i,\ i \in \mathcal{I}\}$ kokioje nors erdvėje $\mathcal{X}$, ir norime rasti taisyklę vadinamą klasifikatoriumi (*classifier*) pagal kurią galėtume kuo tiksliau priskirti klasę[1] $\mathbf{y} \in \mathcal{Y}$ kiekvienam tos erdvės taškui. Kadangi klasių iš anksto nežinome, tai toks uždavinys vadinamas klasterizavimo uždaviniu (*unsupervised learning*). Dažnai pasitaiko, kad galimų klasių skaičius yra iš anksto žinomas (pvz., yra žinoma galimų taškų spalvų aibės galia). Tuomet turime klasterizavimo uždavinį su apribojimu. Neretai žinomos ir dalies stebėjimų (pvz., erdvės taškų) klasės. Tuomet šį uždavinį vadiname dalinai prižiūrimu klasifikavimu (*semi-supervised learning*). Jeigu visų turimų stebėjimų klasės žinomos, tai šį uždavinį vadiname prižiūrimu klasifikavimu (*supervised learning*).

1.2. Statistinis klasifikavimas

Kai aibė $\mathcal{Y}$ sudaryta iš visų klasių kurioms gali priklausyti stebėjimų aibės $\mathcal{X}$ elementai yra žinoma, ir iš baigtinės stebėjimų aibės su žinomomis klasėmis (prižiūrimas klasifikavimas) bandoma daryti išvadas apie visų galimų stebėjimų $\mathbf{x} \in \mathcal{X}$ klases, tai šis uždavinys dažnai vadinamas **statistiniu klasifikavimu**, ir yra formalizuojamas taip:

Tarkime turime stebėjimus $\{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n)\}$. Ieškome klasifikatoriaus $h : \mathcal{X} \to \mathcal{Y}$, kuris bet kurį objektą $\boldsymbol{x} \in \mathcal{X}$ atvaizduoja į jo tikrąją klasę $y \in \mathcal{Y}$, apibrėžtą nežinomo tikrojo

[1] Klasifikavimo uždavinio atveju $\mathcal{Y}$ yra baigtinė, jei taip nėra, tai turime regresijos uždavinį.

atvaizdžio $g : \mathcal{X} \to \mathcal{Y}$.

1.3. Klasifikatoriai

Klasifikavimo metodika dažniausiai remiasi dviem duomenų interpretacijom: statistine ir struktūrine. Statistinis klasifikavimas remiasi statistiniu duomenų apibūdinimu laikant, kad duomenys yra atsitiktinių dydžių reikšmės. Struktūrinis klasifikavimas paremtas struktūriniais požymių sąryšiais, kurie gali būti apibūdinami kintamo kardinalumo simbolių vardinių atributų požymiais[2]. Toliau šiame darbe klasifikavimo klausimą nagrinėju daugiau statistinės nei struktūrinės interpretacijos požiūriu.

Tarkime, kad aibė $X \subseteq \mathcal{X}$ yra sudaryta iš taškų $\{x_i\}_{i=1}^{k}$, čia $k, i \in \mathbb{N}$, kurių kiekvienas yra atsitiktinio vektoriaus $\vec{\xi} : \Omega_{\mathcal{X}} \to \mathbb{R}^n$ realizacija, o aibė $Y \subseteq \mathcal{Y}$ yra sudaryta iš taškų $\{y_i\}_{i=1}^{k}$, kurių kiekvienas yra atsitiktinio dydžio $\varphi : \Omega_{\mathcal{Y}} \to \mathcal{C}$, $|\mathcal{C}| = m < \infty$ priklausomo nuo $\vec{\xi}$ atsitiktinio vektoriaus taip, kad i-tasis Y elementas y_i atitinka i-tajį X elementą, ir reiškia vieną iš $m \in \mathbb{N}$ jo galimų klasių $\mathcal{C}$.

Gerai žinoma, kad jeigu sąlyginiai tankiai $p(x|y_i),\ i = 1, ..., m$ bei apriorinės tikimybės α_i atsitiktinai pasirinkti kiekvieną iš klasių yra žinomos, tai tokio uždavinio optimalų sprendinį duoda Bajeso klasifikatorius

$$h : Y(x) = \arg \max_{1 \leqslant j \leqslant m} \{\alpha_j p(x|y_j)\}, \tag{1.1}$$

nes jis minimizuoja Bajeso klaidą $\sum_{j \neq Y(x)} \alpha_j p(x|y_j)$, kur $Y(x)$ apibrėžtas pagal (1.1) formulę, kurioje funkcija arg įgyja tą j reikšmę, su kuria reiškinys $\alpha_j p(x|y_j)$ įgyja maksimumą.

Yra ir kitokių klasifikatorių, kaip kad sprendimų medžiai, perceptronas, neuroniniai tinklai, naujų, kaip kad atraminių vektorių klasifikatoriai, tačiau Bajeso klasifikatorius yra geriausias

[2]Tai kartais yra patogu, nes leidžia atvaizduoti sudėtingesnius duomenų sąryšius negu įmanoma fiksuoto dimensijų skaičiaus skaitinių požymių vektoriais.

teoriškai įmanomas klasifikatorius.

Visgi, dažnai nustatyti daugiamačių sąlyginių skirstinių tankio funkcijas ir jų parametrus yra sudėtinga, nes didelio matavimo erdvėse tankio funkcijos turi labai didelius parametrų skaičius, be to, norint pakankamai tiksliai įvertinti tankio funkcijas reikia labai didelių imčių. Mėginama taikyti paprastesnius parametrinius klasifikatorius su prielaidomis, kad duomenys yra pasiskirstę pagal kokį nors skirstinį. Tokių klasifikatorių pavyzdžiai - tai tiesinis (Fišerio) klasifikatorius, kvadratinis klasifikatorius, dalimis tiesinis klasifikatorius.

Tačiau net ir tuomet, kai sąlyginių tankių funkcijas pasirenkame iš anksto, dėl daugiadimensiškumo skirstinių parametrų vertinimas dažnai išlieka sunkiai įgyvendinamas, pavyzdžiui, kuomet tankiai yra daugiamačiai Gauso skirstinio, tai tenka skaičiuoti didelę atvirkštinę kovariacijų matricą, jos determinantą.

Egzistuoja ir neparametriniai klasifikavimo metodai kaip kad *k-artimiausių kaimynų* bei *Prazen* metodai. Šie klasifikatoriai tinka ir aukšto dimensijų skaičiaus duomenims, tačiau yra jautrūs jų parametrams, kaip kad kaimynų skaičiui, ir nepasižymi tokiomis geromis teorinėmis savybėmis kaip Bajeso klasifikatorius.

Norint daugelio dimensijų duomenims taikyti Bajeso klasifikatorių bei daugelį kitų klasifikatorių, būtina sumažinti dimensijų skaičių.

1.4. Požymių išskyrimas

Požymių išskyrimas (*feature extraction*) – tai duomenų X atvaizdavimas į kito matavimo erdvę, dažniausiai su tikslu sumažinti erdvės matavimų skaičių. Tą galima padaryti kokiu nors atvaizdžiu $\phi : X \to \mathcal{S}$, kur erdvės $\mathcal{S}$ dimensija yra mažesnė už erdvės $\mathcal{X}$ dimensiją, pavyzdžiui, jei $\mathcal{X} = \mathbb{R}^n$, tai duomenis atvaizduojant į erdvę $\mathcal{S} = \mathbb{R}^d, d \leqslant n$.

Paprasčiausias požymių išskyrimo pavyzdys. Tarkime, turime duomenis $X \subseteq \mathcal{X} \subseteq \mathbb{R}^n$, $n =$

100, tačiau pastebime, kad Y priklauso tik nuo $x^{(15)}$ (pvz., Y įgyja lygiai tas pačias reikšmes kaip $x^{(15)}$), o visos kitos komponentės yra tik triukšmas. Tuomet galime atmesti kitas komponentes, ir klasifikuoti stebėjimus stebėdami tik 15-tą komponentę.

1.5. Požymių išskyrimo reikalingumas

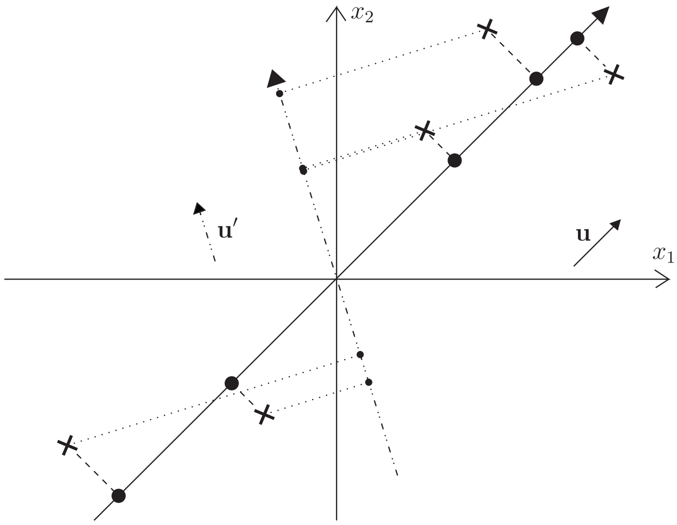
Savo darbo praktikos metu sprendžiau pacientų klasifikavimo į rizikos grupes pagal elektrokardiogramos požymius uždavinį. Šių požymių skaičius buvo 168, ir naudojantis įprastiniais klasifikavimo metodais kaip logistinė regresija ar Fišerio klasifiaktorius, klasifikuoti nepavyko dėl per didelio dimensijų skaičiaus. Požymių išskyrimas pasirodė būtinas, ir tai buvo viena iš priežasčių, kodėl bakalaurinio darbo rašymo metu pasirinkau labiau pasigilinti į naujausią požymių išskyrimo metodiką.

2. ANALITINĖ DALIS

2.1. Pagrindinių komponenčių metodas

Pagrindinių komponenčių analizė yra koordinačių transformacijos metodas 1901 metais sugalvotas Karl Pearson'o. Jis leidžia rasti ortonormuotą bazę, maksimizuojančią turimų erdvės taškų projekcijų dispersijas ortuose, t.y., parenkant ortus taip, kad pradinių taškų projekcijų pirmajame orte dispersija būtų didžiausia įmanoma, tuomet fiksavus pirmąjį ortą antrajame orte projekcijų dispersija būtų didžiausia įmanoma, ir t.t. Koordinačių transformaciją galime rasti maksimizuodami projekcijų dispersijas.

Tarkime turime stebėjimų aibę $X = \{\boldsymbol{x}_i, i = 1, 2, ..., m\} \subset \mathcal{X}$ atitinkančią n-matės Euklidinės erdvės taškus, ir juos norime atvaizduoti į $d < n$ matavimo erdvę pagrindinių komponenčių metodu. Tuomet pirmiausia norime rasti kryptį, kuria projekcijų dispersija didžiausia, todėl ieškome vektoriaus-orto $\mathbf{u}$, kurio kryptimi išvestoje tiesėje suprojektavus pradinius taškus, projekcijų dispersija būtų didžiausia (pav. 1).



1 pav.: *Kryžiukai atitinka aibės X taškus jeigu jie dvimačiai. Juodi taškai rodo jų projekcijas. Projekcijų vektoriaus $\boldsymbol{u}$ krypimi dispersija yra didesnės nei bet kuria kita kryptimi $\boldsymbol{u}'$.*

Kadangi $\mathbf{u}$ yra vienetinis vektorius, tai kito vektoriaus $\boldsymbol{x}_i$ projekciją jame yra $\mathbf{p}_{\mathbf{u}}(\boldsymbol{x}_i) = \boldsymbol{x}_i^T\mathbf{u}$. Jeigu visų projekcijų aibė yra $\{\boldsymbol{x}_i^T\mathbf{u}, i = 1, 2, ..., m\}$, tai dispersijos įvertis yra $\frac{1}{m}\sum_{i=1}^{m}(\boldsymbol{x}_i^T\mathbf{u})^2$, todėl ieškome

$$\max_{\mathbf{u}:\|\mathbf{u}\|=1} \frac{1}{m}\sum_{i=1}^{m}(\boldsymbol{x}_i^T\mathbf{u})^2. \tag{2.1}$$

Kadangi

$$\frac{1}{m}\sum_{i=1}^{m}(\boldsymbol{x}_i^T\mathbf{u})^2 = \frac{1}{m}\sum_{i=1}^{m}(\mathbf{u}^T\boldsymbol{x}_i)(\boldsymbol{x}_i^T\mathbf{u}) = \mathbf{u}^T\left[\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{x}_i\boldsymbol{x}_i^T\right]\mathbf{u} = \mathbf{u}^T\Sigma\mathbf{u}, \tag{2.2}$$

tai ieškome $\max_{\|\mathbf{u}\|=1}\mathbf{u}^T\Sigma\mathbf{u}$. Laikydami, kad norma yra kanoninė skaliarinė sandauga, turime $\mathbf{u}^T\Sigma\mathbf{u}$ maksimizavimą su apribojimu $\mathbf{u}^T\mathbf{u} = 1$, todėl pritaikome Lagranžo daugiklių metodą:

$$\mathcal{L}(\mathbf{u}, \lambda) = \mathbf{u}^T\Sigma\mathbf{u} - \lambda(\mathbf{u}^T\mathbf{u} - 1) \tag{2.3}$$

Diferencijuodami pagal $\mathbf{u}$ ir prilyginę nuliui, gauname:

$$\mathcal{L}'_{\mathbf{u}} = \Sigma\mathbf{u} - \lambda\mathbf{u} = 0 \tag{2.4}$$

Pagal tikrinės reikšmės apibrėžimą tai reiškia, kad maksimizavimo uždavinio sprendinys yra matricos Σ tikrinė tikrinė reikšmė.

Kai $\mathrm{E}X = 0$, tai $\mathrm{cov}(X, X) = \frac{1}{m}\sum_{i=1}^{m} \boldsymbol{x}_i \boldsymbol{x}_i$, todėl Σ yra kovariacijos matrica, ir norint rasti pirmą pagrindinę komponentę, imame kovariacijos matricos didžiausią tikrinę reikšmę, ir randame ją atitinkantį tikrinį vektorių. Pradinę duomenų matricą padauginę iš to vektoriaus gauname duomenų projekcijas pirmojoje ašyje.

Jei norime suprojektuoti duomenis į didesnio dimensijų skaičiaus poerdvį, tai pasirenkame daugiau **u** vektorių atitinkančių daugiau didžiausių tikrinių reikšmių, ir iš jas atitinkančių tikrinių vektorių padauginę pradinę matricą X, gauname duomenų projekcijas į daugiau ortų. Visi tikriniai vektoriai sudaro projekcijų matricą P iš kurios padauginę pradinių duomenų matricą gauname stebėjimus naujose koordinatėse $\tilde{X} = XP$.

Kadangi dažnai kelios ($d \ll n$) komponenčių nusako didelę dalį dispersijos $\sum_{i=1}^{d} D\tilde{\boldsymbol{x}}_i \approx \sum_{i=1}^{n} D\boldsymbol{x}_i$, tai tokiu būdu pagrindinių komponenčių metodas leidžia sumažinti stebėjimų dimensijų skaičių prarandant nedaug informacijos apie dispersiją.

2.1.1. Metodo trūkumai, kovariacijų matrica ir jos alternatyvos

Pagrindinių komponenčių metodas yra labai patogi priemonė matavimų skaičiui sumažinti, tačiau turi savo trūkumų.

- Pagrindinių komponenčių metode nėra galimybės panaudoti informacijos apie stebėjimus žinomomis klasėmis Y klases.
- Kovariacijų matrica Σ yra sudaryta iš kovariacijų. Kovariacija yra tiesinės priklausomybės matas, ir netinka aptikti netiesinei priklausomybei. Norint aptikti netiesinę priklausomybę, būtų gerai turėti matricą Σ sudarytą iš elementų, kurie atspindi ir netiesinę priklausomybę.

Neseniai (Gretton, 2004) [3] sugalvotas netiesinės priklausomybės Hilberto-Šmidto nepriklausomumo kriterijus (HSIC) leidžia aptikti bet kokią netiesinę priklausomybę, todėl "pakeisdami" kovariacijos įverčius šio kriterijaus įverčiais, analogiškai maksimizuodami išraišką $\mathbf{u}^T\Sigma\mathbf{u}$, galime tikėtis jautresnio pagrindinių komponenčių metodo analogo.

Pasirodo, kad neseniai (P. Daniušis, 2008, Y. Zhang 2008) [2],[7] remiantis šiuo kriterijumi buvo išvesti tokie nauji komponenčių analizės metodai, ir jie neturi viršuje minėtų dviejų įprastinio pagrindinių komponenčių metodo trūkumų, ir jie mano praktikos darbe labai tiko, todėl bakalaurinio darbo metu pasirinkau pasidomėti HSIC, juo paremtais metodais, ir pasigilinti į teorija, kuria paremtas šis kriterijus ir juo grindžiami metodai, vadinama Branduolių metodais (*Kernel Methods*).

2.2. Branduolių metodai

Branduolių metodai yra nauja, nuo 1990 metų sparčiai besiplėtojanti duomenų analizės metodika, kuri pradėjo būti taikomi labai įvairiose duomenų analizės srityse. Šių metodų esminis principas remiasi idėja, jog duomenis galime atvaizduoti į saviredukuojančio branduolio Hilberto erdves, jose apibrėžti statistines sąvokas ir algoritmus, ir branduolio funkcijos pagalba juos panaudoti praktikoje. Didžiausias šių metodų privalumas yra tai, kad atvaizdavus duomenis į labai aukšto matavimo erdves dažnai netiesinius uždavinius galime suvesti į tiesinius. Toliau pateikiami pagrindiniai šių metodų taikymą pagrindžiantys teiginiai ir jų įrodymai pagal [1].

Apibrėžimas 1. *Branduoliu* vadiname funkciją $\mathcal{K} : X^2 \to \mathbb{R}$, jei ji tenkina savybes:

1. *Simetriškumo:* $\mathcal{K}(x_1, x_2) = \mathcal{K}(x_2, x_1)$, kai $x_1, x_2 \in X$,

2. *Teigiamo pusapibrėžtumo*[3]: matrica $K = \{\mathcal{K}(x_i, x_j), i, j = 1, ..., T\}$ yra teigiamai pusapi-

[3]Pastaba: statistikoje žodžiu branduolys įprasta vadinti bendresnę funkciją, kuriai nebūtinai galioja teigiamas apibrėžtumas, pavyzdžiui, *Prazen* branduolio atveju $p(x) = \frac{1}{N}\sum_{i=1}^{N} k(x, X_i)$ branduoliu vadinama funkcija

brėžta ($\forall \xi \in \mathbb{R}^n, \xi^T K \xi \geq 0$) visiems $T \in \mathbb{N}$, ir visiems $x_1, x_2, ..., x_T \in X$.

Tarkime H yra Hilberto erdvė tenkinanti savybes:

1. Ji yra vektorinė erdvė virš Realiųjų skaičių skaičių kūno,

2. Joje apibrėžta skaliarinė sandauga $\langle \cdot, \cdot \rangle$,

3. Ji yra pilna metrinė erdvė.

Laikykime, kad H nebūtinai separabili.

Teorema 1. Bet kokiai aibei X funkcija $\mathcal{K} : X^2 \to \mathbb{R}$ yra branduolys tada ir tik tada, kai egzistuoja atvaizdis Φ iš X į Hilberto erdvę $\mathcal{H}$ su skaliarine sandauga $\langle \cdot, \cdot \rangle$, toks, kad visiems $x_1, x_2 \in X$ galioja lygybė $\mathcal{K}(x_1, x_2) = \langle \Phi(x_1), \Phi(x_2) \rangle$.

Būtinumas išplaukia iš branduolio apibrėžimo, nes Gramo matrica yra teigiamai pusapibrėžta ir simetriška. Pakankamumas seka iš toliau įrodysiamų savybių.

Apibrėžimas 2. Tarkime, kad $\mathcal{F}$ yra erdvė sudaryta iš funkcijų apibrėžtų aibėje X. Funkcija $\mathcal{K}(x_1, x_2)$ yra *saviredukuojantis branduolys* (s.b.) atitinkantis erdvę $\mathcal{F}$ jei:

1. Kiekvienam $x \in X$, funkcija $\mathcal{K}(x, \cdot)$ priklauso Hilberto erdvei $\mathcal{F}$,

2. Galioja *saviredukuotumo savybė*: kiekvienam $f \in \mathcal{F}$ ir kiekvienam $x \in X$ galioja lygybė $f(x) = \langle f, \mathcal{K}(x, \cdot) \rangle$.

Sąvoka: Erdvė $\mathcal{F}$ su saviredukuojančiu branduoliu yra vadinama *saviredukuojančio branduolio Hilberto erdve (SBHE).* Angliškai: Reproducing Kernel Hilbert Space (RKHS).

Saviredukuojančių branduolių savybės

Teorema 2.

1. Savybė: Jei erdvei $\mathcal{F}$ egzistuoja s.b., tai jis vienintelis.

$k(x_1, x_2)$ nebūtinai teigiamai apibrėžta.

2. Savybė: Jei $\mathcal{K}$ yra s.b. atitinkantis erdvę $\mathcal{F}$, tai visiems $x \in X$ bei $f \in \mathcal{F}$ galioja nelygybė $|f(x)| \leq \sqrt{\mathcal{K}(x,x)} \|f\|_{\mathcal{F}}$.

3. Savybė: Jei $\mathcal{F}$ yra SBHE, tai iš konvergavimo erdvėje $\mathcal{F}$ išplaukia pataškis atitinkamų funkcijų (kurios yra $\mathcal{F}$ elementai) konvergavimas.

Įrodymas:

$\bigtriangledown$

Tarkime, kad yra du s.b. $\mathcal{K}_1$ ir $\mathcal{K}_2$ kokiai nors vienai erdvei $\mathcal{F}$. Tuomet kiekvienam $x \in X$ funkcija $\mathcal{K}_1(x,\cdot) - \mathcal{K}_2(x,\cdot)$ priklauso $\mathcal{F}$, ir taikydami branduolio tiesiškumo ir saviredukuotumo savybes, gauname, kad

$$\begin{aligned}
\|\mathcal{K}_1(x,\cdot) - \mathcal{K}_2(x,\cdot)\|_{\mathcal{F}}^2 &= \langle \mathcal{K}_1(x,\cdot) - \mathcal{K}_2(x,\cdot), \mathcal{K}_1(x,\cdot) - \mathcal{K}_2(x,\cdot)\rangle \\
&= \langle \mathcal{K}_1(x,\cdot) - \mathcal{K}_2(x,\cdot), \mathcal{K}_1(x,\cdot)\rangle - \\
&\qquad \langle \mathcal{K}_1(x,\cdot) - \mathcal{K}_2(x,\cdot), \mathcal{K}_2(x,\cdot)\rangle \\
&= (\mathcal{K}_1(x,x) - \mathcal{K}_2(x,x)) - (\mathcal{K}_1(x,x) - \mathcal{K}_2(x,x)) \\
&= 0.
\end{aligned}$$

Iš Hilberto erdvės apibrėžimo išplaukia, kad $\mathcal{K}_1(x,\cdot)$ ir $\mathcal{K}_2(x,\cdot)$ sutampa, ir todėl kaip funkcijos jie visur lygūs.

(2) savybė seka iš saviredukuotumo savybės ir Koši-Buniakovo nelygybės.

(3) savybė seka iš (2), nes visiems $f_1, f_2 \in \mathcal{F}$ ir $x \in X$ galioja nelygybė

$$|f_1(x) - f_2(x)| \leq \sqrt{\mathcal{K}(x,x)} \|f_1 - f_2\|_{\mathcal{F}}.$$

$\bigtriangleup$

Sąvoka: Tarkime, kad $\mathcal{F}$ yra Hilberto erdvė sudaryta iš realiųjų funkcijų apibrėžtų aibėje X. Imkime $x \in X$, ir sudarykime funkcionalą $\mathcal{F} \to \mathbb{R}$ priskiriantį funkcijai f reikšmę

$f(x)$. Šis funkcionalas yra tiesinis (f požiūriu), ir vadinamas *vertinimo funkcionalu* (evaluation functional).

Teorema 3. Hilberto erdvė $\mathcal{F}$ sudaryta iš realiųjų funkcijų apibrėžtų aibėje X yra SBHE tada ir tik tada, kai kiekvienam $x \in X$ atitinkamas vertinimo funkcionalas yra tolydus.

Įrodymas:

∇

Pakankamumas išplaukia iš antros saviredukuojančių branduolių savybės.

Būtinumui įrodyti pasiremiame Ryso-Fišerio reprezentacijos teorema, kuri sako, kad kiekvienas tolydus tiesinis funkcionalas Hilberto erdvėje gali būti užrašytas kaip skaliarinė sandauga iš tos erdvės elemento.

Imkime $x \in X$. Kadangi vertinimo funkcionalas tolydus, tai iš Ryso-Fišerio reprezentacijos teoremos išplaukia, kad egzistuoja vienintelis branduolys $k_x \in \mathcal{F}$ tenkinantis lygybę $f(x) = \langle f, k_x \rangle$. Naudodamiesi šiuo branduoliu apibrėžkime atvaizdį $F : X \to \mathcal{F}$ tokį, kad $F(x) = k_x$. Tarkime, kad $\mathcal{K}(x_1, x_2) = \langle F(x_1), F(x_2) \rangle$.

Tuomet branduolį galime užsirašyti kaip x_2 funkciją

$$\begin{aligned} \mathcal{K}(x_1, x_2) &= \langle k_{x_1}, k_{x_2} \rangle \\ &= k_{x_1}(x_2), \end{aligned}$$

be to, $\mathcal{K}(x_1, \cdot) = k_{x_1}(\cdot) \in \mathcal{F}$. Tokiu būdu iš Ryso-Fišerio teoremos išplaukia, kad kiekvienam $f \in \mathcal{F}$ ir $x \in X$

$$\begin{aligned} f(x) &= \langle f, k_x \rangle \\ &= \langle f, \mathcal{K}(x, \cdot) \rangle. \end{aligned}$$

Taigi, $\mathcal{K}$ yra s.b. atitinkantis erdvę $\mathcal{F}$.

$\triangle$

Ši tapatybė svarbi tuo, kad vertinimo funkcionalo tolydumas reiškia, jog jeigu funkcijos f_1 ir f_2 yra arti viena kitos Hilberto erdvės normos prasme, yra atvaizduojamos į reikšmes $f_1(x)$ ir $f_2(x)$, kurios yra arti viena kitos.

Taigi, parodėme, kad s.b. $\mathcal{K}(x_1, x_2)$ gali būti užrašytas kaip $\langle F(x_1), F(x_2)\rangle$, ir tai reiškia, kad $\mathcal{K}$ yra simetriškas (dėl skaliarinės sandaugos simetriškumo), ir teigiamai pusapibrėžtas (t.y., iš jo reikšmių gaunamų kai $x_1, x_2, \ldots, x_T \in X$ sudaryta matrica yra teigiamai pusapibrėžta).

Toliau formuluojama teorema parodo, kad branduolių ir saviredukuojančių branduolių klasės sutampa.

Teorema 4. (Egzistavimo teorema) Tarkime, kad $\mathcal{K} : X^2 \to \mathbb{R}$ yra reali dviejų argumentų funkcija apibrėžta aibėje X. Tuomet $\mathcal{K}$ yra saviredukuojantis branduolys kokiai nors Hilberto erdvei sudarytai iš funkcijų $\mathcal{F}$ apibrėžtų aibėje X tada ir tik tada, kai

- $\mathcal{K}$ yra simetriška
- $\mathcal{K}$ yra teigiamai pusapibrėžta.

(Jei yra erdvė turi branduolį $\mathcal{K}$, tai ji vienintelė.)

Įrodymas ilgas: pirmiausia sudaroma erdvė $\mathcal{F}_1$ iš branduolio tiesinių daugdarų ir parodoma, kad jos sudaro erdvę su skaliarine sandauga. Tuomet parodoma, kad ji turi papildinį (šiuo atveju ieškome papildinio sudaryto iš $X \to \mathbb{R}$ funkcijų, ir metrinės erdvės papildinio nepakanka), ir sudaroma pilnoji erdvė $\mathcal{F}$. Tuomet parodomas vienatinumas: kad vienam branduoliui $\mathcal{K}$ egzistuoja tik viena saviredukuojančio branduolio Hilberto erdvė $\mathcal{F}$.

Įrodymas:

$\triangledown$

Tarkime $\mathcal{K} : X^2 \to \mathbb{R}$ yra branduolys.

3.3.1. Tiesinė erdvė su skaliarine sandauga

Sudarykime tiesinę erdvę iš $\mathcal{F}_1$ iš tiesinių daugdarų $\sum_{i=1}^n a_i\mathcal{K}(x_i,\cdot)$, kur $n \in \mathbb{N}$, $a_i \in \mathbb{R}$, $x_i \in X$. Tiesiškumas seka iš to, kad daugdaros yra tiesinės.

Apibrėžkime skaliarinę sandaugą kaip operaciją apibrėžiamą lygybe

$$\left\langle \sum_{i=1}^n a_i\mathcal{K}(x_i,\cdot), \sum_{i=1}^n b_i\mathcal{K}(x_i,\cdot) \right\rangle = \sum_{i,j=1}^n a_i b_j \mathcal{K}(x_i,x_j)$$

(pridėdami narių su nuliniais koeficientais galėtume užtikrinti, kad tiesinių daugdarų narių skaičius ir eiliškumas sutaptų.)

Kadangi bandome sudaryti funkcijų erdvę, tai pirmiausia turime parodyti, kad taip apibrėžta skaliarinė sandauga yra nepriklausoma nuo dauginamųjų funkcijų išraiškos.

Tarkime, kad $f(x) = \sum_{i=1}^n a_i\mathcal{K}(x_i,x)$ ir $g(x) = \sum_{i=1}^n b_i\mathcal{K}(x_i,x)$. Tuomet

$$\begin{aligned}\langle f,g\rangle &= \sum_{i,j=1}^n a_i b_j \mathcal{K}(x_i,x_j)\\ &= \sum_{i=1}^n a_i \left(\sum_{j=1}^n b_j\mathcal{K}(x_i,x_j)\right)\\ &= \sum_{i=1}^n a_i g(x_j).\end{aligned}$$

Matome, kad skaliarinė sandauga gali būti išreikšta per funkcijos g reikšmes, ir todėl nepriklausoma nuo konkretaus funkcijos g kaip tiesinės daugdaros išraiškos. Analogiškas teiginys galioja f, ir todėl taip apibrėžta skaliarinė sandauga nepriklauso nuo dauginamųjų funkcijų išraiškų.

Taip apibrėžtos operacijos $\langle\cdot,\cdot\rangle$ simetriškumas seka iš to, kad branduolys $\mathcal{K}$ yra simetriškas. Be to, imdami viršuje minėtą f, turime

$$\langle f,f\rangle = \sum_{i,j=1}^n a_i a_j \mathcal{K}(x_i,x_j) \geq 0,$$

nes $\mathcal{K}$ yra teigiamai pusapibrėžtas. Taigi, ir operacija $\langle\cdot,\cdot\rangle$ yra teigiamai pusapibrėžta, ir tokiu būdu parodėme, kad operacija $\langle\cdot,\cdot\rangle$ turi teigiamai pusapibrėžtą simetrinę bitiesinę (abiejų argumentų požiūriu tiesinę) formą. Lieka parodyti, kad ji yra teigiamai apibrėžta, ir todėl yra skaliarinė sandauga.

Tuo tikslu įvertinkime $\langle f(\cdot), \mathcal{K}(x, \cdot)\rangle$, kur $f \in \mathcal{F}_1$ ir $x \in X$. Gauname, kad

$$\langle f(\cdot), \mathcal{K}(x, \cdot)\rangle = f(x),$$

ir todėl operacija $\langle \cdot, \cdot \rangle$ bei branduolys $\mathcal{K}$ tenkina saviredukuotumo savybę.

Kadangi operacija $\langle \cdot, \cdot \rangle$ yra teigiamai pusapibrėžta, tai jai galioja Koši-Buniakovo nelygybė, iš kurios seka, kad

$$\langle f, g \rangle \leq \|f\| \cdot \|g\|,$$

čia $\|f\|$ yra apibrėžiama kaip $\sqrt{\langle f, f \rangle}$. Tuomet iš ką tik įrodytos saviredukuotumo savybės seka, kad

$$\begin{aligned} \langle f(\cdot), \mathcal{K}(x, \cdot)\rangle &\leq \|f\| \cdot \|\mathcal{K}(x, \cdot)\| \\ &= \|f\| \sqrt{\mathcal{K}(x, x)}. \end{aligned}$$

Vadinasi iš to, kad $\|f\| = 0$ išplaukia, jog $f(x) = 0$ visiems $x \in X$, ir tai reiškia, kad operacija $\langle \cdot, \cdot \rangle$ yra teigiamai apibrėžta, ir todėl iš tikrųjų yra skaliarinė sandauga.

3.3.2. Pilnumas

Parodysime, kad $\mathcal{F}_1$ yra pilnoji erdvė.

Tarkime, kad $f_1, f_2, \ldots \in \mathcal{F}_1$ yra fundamentalioji seka. Tuomet iš viršuje gautos nelygybės seka, kad kiekvienam $x \in X$ galioja nelygybė

$$\begin{aligned} |f_n(x) - f_m(x)| &= |\langle f_n - f_m, \mathcal{K}(x, \cdot)\rangle| \\ &\leq \|f_n - f_m\| \sqrt{\mathcal{K}(x, x)}, \end{aligned}$$

iš kurios seka, kad $f_1(x), f_2(x), \ldots$ yra fundamentalioji seka, ir turi ribą. Tuomet galime apibrėžti funkciją $f : X \to \mathbb{R}$ tokią, kad $f(x) = \lim_{n\to\infty} f_n(x)$.

Tarkime, kad $\mathcal{F}$ yra sudaryta iš visų taip apibrėžtų funkcijų. Matome, kad $\mathcal{F}_1 \subseteq \mathcal{F}$, kadangi kiekvieną f galime užrašyti kaip $\mathcal{F}_1$ elementą, kuris yra pataškė sekos $f, f, \ldots$ riba.

Tuomet skaliarinę sandaugą erdvėje $\mathcal{F}$ irgi galime apibrėžti: jei f yra sekos $f_1, f_2, \ldots \in \mathcal{F}_1$ pataškė riba, o g yra sekos $g_1, g_2, \ldots \in \mathcal{F}_1$ pataškė riba, tai $\langle f, g\rangle_{\mathcal{F}} = \lim_{n\to\infty}\langle f_n, g_n\rangle_{\mathcal{F}_1}$.

Ši riba egzistuoja, nes visiems natūriniams n_1, n_2, m_1 ir m_2 galioja nelygybė

$$
\begin{aligned}
&|\langle f_{n_1}, g_{m_1}\rangle - \langle f_{n_2}, g_{m_2}\rangle| \leq \\
&\qquad |\langle f_{n_1}, g_{m_1}\rangle - \langle f_{n_1}, g_{m_2}\rangle| + |\langle f_{n_1}, g_{m_2}\rangle - \langle f_{n_2}, g_{m_2}\rangle| \leq \\
&\qquad\qquad \|f_{n_1}\| \cdot \|g_{m_1} - g_{m_2}\| + \|g_{m_2}\| \cdot \|f_{n_1} - f_{n_2}\|.
\end{aligned}
$$

Kadangi fundamentaliosios sekos elementų normos yra tolygiai aprėžtos, tai šis skirtumas esant pakankamai dideliem n_1, n_2, m_1 ir m_2 gali būti kiek norima mažas. Taigi, kadangi visiems pakankamai dideliems n ir m, skirtumas $|\langle f_n, g_m\rangle - s|$ tampa kiek norima mažas, tai egzistuoja netgi kelių kintamųjų riba $\lim_{n,m\to\infty}\langle f_n, g_m\rangle = s$.

Be to, apibrėžtoji skaliarinė sandauga egzistuoja nepriklausomai nuo pasirinktų fundamentaliųjų sekų konverguojančių į f ir g. Tarkime, kad turime dvi fundamentaliųjų sekų poras $f_1, f_2, \ldots$ ir $f'_1, f'_2, \ldots$ konverguojančias į f, bei $g_1, g_2, \ldots$ ir $g'_1, g'_2, \ldots$ konverguojančias į g. Nagrinėkime išraišką $\langle f_m - f'_m, g_n\rangle$. Seka sudaryta iš funkcijų $f_n - f'_n$ irgi yra fundamentalioji, ir todėl turi egzistuoti riba $\lim_{n,m\to\infty}\langle f_m - f'_m, g_n\rangle$. Įvertinkime šią ribą. Egzistuoja koeficientai $b_1^m, b_2^m, \ldots, b_p^m$ ir elementai $z_1^m, z_2^m, \ldots, z_p^m$, tokie, kad $f(\cdot) = \sum_{i=1}^p b_i^m \mathcal{K}(z_i^m, \cdot)$ ($p = p(m)$ gali įgyti skirtingas reikšmes priklausomai nuo m), todėl galime užrašyti

$$
\begin{aligned}
\langle f_m - f'_m, g_n\rangle &= \langle f_m - f'_m, \sum_{i=1}^p b_i^m \mathcal{K}(z_i^m, \cdot)\rangle \\
&= \sum_{i=1}^m b_i^m (f_m(z_i) - f'_m(z_i)).
\end{aligned}
$$

Kadangi f_m ir f'_m pataškiui konverguoja į 0, tai šis reiškinys artėja į nulį, kai $m \to \infty$. Iš to seka, kad

$$\lim_{n,m\to\infty} \langle f_m - f'_m, g_n \rangle = 0.$$

Analogiškai,

$$\lim_{n,m\to\infty} \langle f'_m, g_n - g'_n \rangle = 0.$$

Taigi, skirtumas

$$\langle f_m, g_n \rangle - \langle f'_m, g'_n \rangle = \langle f_m - f'_m, g_n \rangle + \langle f'_m, g_n - g'_n \rangle$$

artėja į 0, kai $n, m \to \infty$, ir todėl skaliarinės sandaugos apibrėžimas nepriklauso nuo konkrečios pasirinktos fundamentaliosios sekos.

Tai, kad operacija $\langle \cdot, \cdot \rangle$ abiejų argumentų požiūriu tiesiška, išplaukia iš jos apibrėžimo ir ribos tiesiškumo (sumos riba yra ribų suma, sandaugos riba yra ribų sandauga). Be to, skaičius $\|f\| = \langle f, f \rangle$ yra neneigiamas kaip neneigiamų skaičių riba. T.y., paėmę fundamentaliąją seką $f_1, f_2, \ldots \in \mathcal{F}_1$ pataškiui konverguojančią į f, turime

$$\begin{aligned} |f(x)| &= \lim_{n\to\infty} |f_n(x)| \\ &\leq \lim_{n\to\infty} \sqrt{\mathcal{K}(x,x)} \|f_n\| \\ &= \sqrt{\mathcal{K}(x,x)} \|f\|, \end{aligned}$$

ir iš $\|f\|$ lygybės išplaukia, kad $f(x) = 0$ visiems $x \in X$.

Parodėme, kad $\mathcal{F}$ iš tikrųjų yra tiesinė erdvė su skaliarine sandauga. Akivaizdu, kad $\mathcal{F}_1 \subseteq \mathcal{F}$, ir skaliarinė sandauga erdvėje $\mathcal{F}$ praplečia skaliarinės sandaugos erdvėje $\mathcal{F}_1$ apibrėžimą.

Dabar parodysime, kad $\mathcal{F}$ yra pilnoji. Pirmiausia tarkime, kad $f_1, f_2 \ldots$ yra erdvės $\mathcal{F}_1$ elementų fundamentalioji seka pataškiui konverguojanti į f. Tuomet

$$\begin{aligned} \|f - f_n\| &= \sqrt{\langle f - f_n, f - f_n \rangle} \\ &= \sqrt{\lim_{m\to\infty} \langle f_m - f_n, f_m - f_n \rangle} \\ &= \sqrt{\lim_{m\to\infty} \|f_m - f_n\|}. \end{aligned}$$

Ši riba konverguoja į 0 kai $n \to 0$, ir todėl f yra sekos f_n riba esanti erdvėje $\mathcal{F}$. Tuomet nagrinėkime erdvės $\mathcal{F}$ elementų seką $f_1, f_2 \ldots$. Kiekvienam n egzistuoja $g_n \in \mathcal{F}_1$ tokia, kad $\|f_n - g_n\| \leq 1/2^n$. Seka $g_1, g_2, \ldots$ yra erdvės $\mathcal{F}_1$ fundamentalioji seka, ir todėl turi ribą erdvėje $\mathcal{F}$. Iš čia seka, kad ši riba sutampa su $f_1, f_2 \ldots$ riba.

Lieka parodyti, kad erdvės $\mathcal{F}$ atžvilgiu galioja saviredukuotumo savybė, o tai seka iš tolydumo. Tarkime, kad $f_1, f_2 \ldots$ yra erdvės $\mathcal{F}_1$ fundamentalioji seka pataškiui konverguojanti į f. Tuomet

$$
\begin{aligned}
f(x) &= \lim_{n\to\infty} f_n(x) \\
&= \lim_{n\to\infty} \langle f_n(\cdot), \mathcal{K}(x, \cdot) \rangle \\
&= \langle f(\cdot), \mathcal{K}(x, \cdot) \rangle
\end{aligned}
$$

Tokiu būdu branduoliui $\mathcal{K}$ gavome SBHE. Pastebėtina, kad pirmoje įrodymo dalyje apibrėžta $\mathcal{F}_1$ yra tiršta erdvėje $\mathcal{F}$.

3.3.3. Vienatis

Parodysime, kad konkrečiam branduoliui $\mathcal{K}$ egzistuoja vieinintelė SBHE. Tarkime, kad $\mathcal{F}$ yra viršuje apibrėžta SBHE, o $\mathcal{H}$ yra kažkokia kita SBHE apibrėžta tam pačiam branduoliui $\mathcal{K}$.

Iš SBHE apibrėžimo išplaukia, kad visos funkcijos $\mathcal{K}(x, \cdot)$ turi priklausyti $\mathcal{H}$. Be to, visos tiesinės daugdaros $\sum_{i=1}^{n} a_i \mathcal{K}(x_i, \cdot)$ taip pat turi priklausyti $\mathcal{H}$. Taigi, šioms erdvėms kaip aibėms galioja sąryšis $\mathcal{F}_1 \subseteq \mathcal{H}$.

Kadangi erdvei $\mathcal{H}$ galioja saviredukuotumo savybė, tai erdvės $\mathcal{F}_1$ elementams skaliarinė sandauga $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ sutampa su anksčiau apibrėžta skaliarine sandauga. Taigi, $\mathcal{F}_1$ yra $\mathcal{H}$ poerdvis.

Kadangi $\mathcal{H}$ yra pilnoji, tai visos fundamentaliosios sekos apibrėžtos erdvėje $\mathcal{F}_1$ turi ribas erdvėje $\mathcal{H}$. Kadangi SBH erdvėse iš konvergavimo išplaukia pataškis konvergavimas, tai visų fundamentaliųjų erdvės $\mathcal{F}_1$ sekų visos pataškės ribos priklauso $\mathcal{H}$. Taigi, kaip aibės, ir $\mathcal{F} \subseteq \mathcal{H}$.

Kadangi skaliarinė sandauga yra savo pačios požiūriu tolydi, tai lygybė

$$\lim_{n\to\infty} \langle f_n, g_n \rangle = \langle f, g \rangle$$

galioja visoms sekoms $f_1, f_2, \dots$ ir $g_1, g_2, \dots$ tokioms, kad $f_n \to f$ ir $g_n \to g$ priklauso $\mathcal{H}$, kai $n \to \infty$. Taigi, skaliarinė sandauga erdvėje $\mathcal{F}$ sutampa su skaliarine sandauga erdvėje $\mathcal{H}$ ($\mathcal{F}$ yra uždaras $\mathcal{H}$ poerdvis).

Tarkime, kad $h \in \mathcal{H}$. Tuomet galime perrašyti $h = h_{\mathcal{F}} + h^{\perp}$, kur $h_{\mathcal{F}} \in \mathcal{F}$, o $h^{\perp}$ yra statmenas erdvei $\mathcal{F}$, ir todėl statmenas visoms funkcijoms $\mathcal{K}(x, \cdot)$ priklausančioms $\mathcal{F}_1 \subseteq \mathcal{F}$. Kadangi saviredukuotumo savybė galioja erdvei $\mathcal{H}$, tai

$$\begin{aligned} h(x) &= \langle h, \mathcal{K}(x, \cdot) \rangle \\ &= \langle h_{\mathcal{F}}, \mathcal{K}(x, \cdot) \rangle + \langle h^{\perp}, \mathcal{K}(x, \cdot) \rangle \\ &= \langle h_{\mathcal{F}}, \mathcal{K}(x, \cdot) \rangle \\ &= h_{\mathcal{F}}(x). \end{aligned}$$

Iš čia išplaukia, kad h sutampa su $h_{\mathcal{F}}$ visur aibėje X, ir $\mathcal{H} = \mathcal{F}$.

$\triangle$

2.3. Hilberto-Šmidto nepriklausomumo kriterijus (HSIC)

Tarkime, kad turime saviredukuojančio branduolio Hilberto erdvę $\mathcal{F}$ sudarytą iš funkcijų apibrėžtų aibėje $\mathcal{X}$ virš realiųjų skaičių kūno $\mathbb{R}$. Pažymėkime $\delta_x : \mathcal{F} \to \mathbb{R}$ – kiekvienam $x \in \mathcal{X}$ aprėžtą tiesinį vertinimo funkcionalą funkcijoms $f \in \mathcal{F}$ priskiriantį reikšmes $f(x) \in \mathbb{R}$. Tuomet kiekvienam taškui $x \in \mathcal{X}$ egzistuoja $\phi(x) \in \mathcal{F}$ toks, kad $\langle \phi(x), \phi(x') \rangle_{\mathcal{F}} = k(x, x')$, kur $k : \mathcal{X}^2 \to \mathbb{R}$ yra vienintelis teigiamai apibrėžtas branduolys. Laikysime, kad $\mathcal{F}$ yra separabili (ji turi turėti pilną ortonormuotą bazę). Anot (Hein ir Beousquet, 2004) tą galima užtikrinti parenkant tolydų branduolį virš separabilios $\mathcal{X}$ (pvz., $\mathbb{R}^n$). Analogiškai apibrėžiame dar vieną separabilią SBHE $\mathcal{G}$ susietą su branduoliu $l(\cdot,\cdot)$ ir į ją duomenis atvaizduojančiu atvaizdžiu ψ virš separabilios $\mathcal{Y}$.

Apibrėžimas 1. Pažymėkime $C : \mathcal{G} \to \mathcal{F}$ – tiesinį operatorių. Tuomet *Hilbert-Šmidto norma* C yra apibrėžiama kaip suma

$$\|C\|_{HS}^2 = \sum_{i,j} \langle Cv_i, u_j \rangle_{\mathcal{F}}^2,$$

jei tik ši suma konverguoja. Čia v_i ir u_j yra atitinkamai erdvių $\mathcal{F}$ ir $\mathcal{G}$ ortonormalios bazės. Hilberto-Šmidto norma yra matricos kvadratinės normos $\|M\|_F^2 = \sum_{i,j=1}^n m_{ij}^2 = \mathrm{tr}(M^T M)$ apibendrinimas.

Apibrėžimas 2. Tiesinis operatorius $C : \mathcal{G} \to \mathcal{F}$ vadinamas *Hilberto-Šmidto operatoriumi* jei egzistuoja jo Hilberto-Šmidto norma. Hilberto-Šmidto operatorių aibė $HS(\mathcal{F}, \mathcal{G}) : \mathcal{G} \to \mathcal{F}$ yra separabilioji Hilberto erdvė su skaliarine sandauga

$$\langle C, D \rangle_{HS} := \sum_{i,j} \langle Cv_i, u_j \rangle_{\mathcal{F}} \langle Dv_i, u_j \rangle_{\mathcal{F}}$$

Apibrėžimas 3. Tarkime, kad $f \in \mathcal{F}$ ir $g \in \mathcal{G}$. Tuomet *tenzorinės sandaugos operatorius* $f \otimes g : \mathcal{G} \to \mathcal{F}$ yra apibrėžiamas lygybe

$$(f \otimes g)h := f \langle g, h \rangle_{\mathcal{G}}, \text{ visiems } h \in \mathcal{G}$$

Apibrėžimas 4. Tarkime, kad $(\mathcal{X}, \Gamma)$ ir $(\mathcal{Y}, \Lambda)$ yra susieti su tikimybiniais matais atitinkamai p_x ir p_y (Γ yra Borelio aibės generuotos $\mathcal{X}$, Λ yra Borelio aibės generuotos $\mathcal{Y}$). Šių matų požiūriu galime apibrėžti *vidurkio elementus* kaip erdvių $\mathcal{F}$ ir $\mathcal{G}$ elementus, kuriems galioja lygybės

$$\langle \mu_x, f \rangle_{\mathcal{F}} := \mathbf{E}_x[\langle \phi(x), f \rangle_{\mathcal{F}}] = \mathbf{E}_x[f(x)],$$

$$\langle \mu_y, g \rangle_{\mathcal{G}} := \mathbf{E}_y[\langle \psi(y), g \rangle_{\mathcal{G}}] = \mathbf{E}_y[g(y)].$$

čia ϕ yra atvaizdis iš $\mathcal{X}$ į SBHE $\mathcal{F}$, vadinamas požymių atvaizdžiu, bei ψ taip pat požymių atvaizdis iš $\mathcal{Y}$ į $\mathcal{G}$. Tuomet norma $\|\mu_x\|_{\mathcal{F}}^2$ gali būti apskaičiuota dukart skaičiuojant vidurkį, ir per branduolio funkciją užrašyta sekančiu būdu

$$\|\mu_x\|_{\mathcal{F}}^2 = \mathbf{E}_{x,x'}[\langle \phi(x), \phi(x') \rangle_{\mathcal{F}}] = \mathbf{E}_{x,x'}[k(x,x')].$$

Čia matematinė viltis su nepriklausomomis x, x' reikšmėmis iš p_x. Vidurkiai μ_x, μ_y egzistuoja, jei tik jų atitinkamos normos erdvėse $\mathcal{F}$ ir $\mathcal{G}$ yra aprėžtos, o taip yra jei branduoliai k ir l yra aprėžti (kadangi tuomet $\mathbf{E}_{x,x'}[k(x,x')] < \infty$ ir $\mathbf{E}_{y,y'}[l(y,y')] < \infty$).

Apibrėžimas 5. (Pagal Baker [1973], Fukumizu [2004]), *kross-kovariacijos operatorius*, susietas su jungtiniu matu p_{xy} apibrėžtu mačioje erdvėje $(\mathcal{X} \times \mathcal{Y}, \Gamma \times \Lambda)$ yra tiesinis operatorius $C_{xy} : \mathcal{G} \to \mathcal{F}$ apibrėžtas lygybe

$$C_{xy} := \mathbf{E}[(\phi(x) - \mu_x) \otimes (\psi(y) - \mu_y)] = \mathbf{E}_{x,y}[\phi(x) \otimes \psi(y)] - \mu_x \otimes \mu_y.$$

Čia antroji lygybė seka iš matematinės vilties tiesiškumo.

Apibrėžimas 6. (Hilberto-Šmidto nepriklausomumo kriterijus) Tarkime, kad $\mathcal{F}$ ir $\mathcal{G}$ yra separabilios SBHE, o p_{xy} yra jungtinis matas mačioje erdvėje $(\mathcal{X} \times \mathcal{Y}, \Gamma \times \Lambda)$, tuomet Hilberto-Šmidto nepriklausomumo kriterijus (HSIC – angl. *Hilbert-Schmidt Independence Criterion*) apibrėžiamas kaip susijusio kross-kovariacijos operatoriaus C_{xy} kvadratinė Hilbert-Šmidto norma:

$$HSIC(p_{xy}, \mathcal{F}, \mathcal{G}) := \|C_{xy}\|_{HS}^2. \tag{2.5}$$

Norint apskaičiuoti HSIC, galima jį išreikšti jį per branduolio funkcijas. Tai pasiekiama sekančiu teiginiu (Gretton, 2004):

$$HSIC(p_{xy}, \mathcal{F}, \mathcal{G}) = \mathbf{E}_{x,x',y,y'}[k(x,x')l(y,y')] + \mathbf{E}_{x,x'}[k(x,x')]E_{y,y'}[l(y,y')] -$$

$$2\mathbf{E}_{x,y}[\mathbf{E}_{x'}[k(x,x')]\mathbf{E}_{y'}[l(y,y')]].$$

Čia $\mathbf{E}_{x,x',y,y'}$ žymi vidurkį skaičiuojamą pagal nepriklausomas poras (x,y) ir (x',y') iš p_{xy}. Šio teiginio įrodymas yra [3]. Ten pat parodoma, kad ši norma egzistuoja tuomet, kai apibrėžime esamos matematinės viltys yra aprėžtos, ir kad taip yra tuomet, kai branduoliai k ir l yra aprėžti.

2.4. Hilberto-Šmidto nepriklausomumo kriterijaus įverčiai $HSIC_0$ ir $HSIC_1$

Norint pritaikyti šį kriterijų praktiškai nepriklausomumui tikrinti, būtina įvertinti HSIC baigtinės imties atveju, parodyti, kad įvertis pakankamai greitai artėja į teorinį HSIC, ir kad HSIC tikrai rodo nepriklausomumą. Ten pat [3] išvedamas empirinis HSIC įvertis, ir pateikiami šie įrodymai. Čia pateikiame tik šio kriterijaus įverčius.

Apibrėžimas 1 (Empirinis HSIC). Tarkime, kad $Z := \{(x_1, y_1), ..., (x_m, y_m)\} \subseteq \mathcal{X} \times \mathcal{Y}$ yra m nepriklausomų stebėjimų generuotų pagal tikimybinį matą p_{xy}. HSIC įvertis $HSIC_0(Z, \mathcal{F}, \mathcal{G})$ apibrėžiamas lygybe

$$HSIC_0(Z, \mathcal{F}, \mathcal{G}) := (m-1)^{-2}\mathrm{tr}KHLH, \tag{2.6}$$

čia $H, K, L \in \mathbb{R}^{m \times m}$, $K_{ij} := k(x_i, x_j)$, $L_{ij} := l(y_i, y_j)$, ir $H_{ij} := \delta_{ij} - m^{-1}$. Šio įverčio privalumas yra tai, kad jo skaičiavimui užtenka atlikti $O(m^2)$ operacijų. Šis įvertis yra paslinktas su poslinkiu $O(m^{-1})$. Šie teiginiai taip pat įrodomi [3].

Panaudojant šį įvertį [7] išvedamas nauja pagrindinių komponenčių metodo modifikacija. Kadangi į šio įverčio skaičiavimą įeina $\mathcal{Y}$ branduolio matrica L, tai gaunamas metodas tinka įtraukti informacijai apie stebėjimus su žinomomis klasėmis, be to, metodas tampa jautrus netiesinei priklausomybei.

Apibrėžimas 2 (Nepaslinktas empirinis HSIC). Tarkime, kad turime tas pačias prielaidas kaip ir pirmąjame apibrėžime šiame skyrelyje, turime tuos pačius m nepriklausomų stebėjimų iš p_{xy}. Tuomet nepaslinktas HSIC įvertis $HSIC_1(Z, \mathcal{F}, \mathcal{G})$ apibrėžiamas

$$HSIC_1(Z, \mathcal{F}, \mathcal{G}) := \frac{1}{m(m-3)}[\mathrm{tr}\tilde{K}\tilde{L} + \frac{\mathbf{1}^T\tilde{K}\mathbf{1}\mathbf{1}^T\tilde{L}\mathbf{1}}{(m-1)(m-2)} - \frac{2}{m-2}\mathbf{1}^T\tilde{K}\tilde{L}\mathbf{1}], \tag{2.7}$$

čia $\tilde{K}$ ir $\tilde{L}$ yra branduolio matricos, kurių įstrižaininiai elementai lygūs nuliams ($\tilde{K} = K - \mathrm{diag}(K), \tilde{L} = L - \mathrm{diag}(L)$).

Panaudojant šį įvertį [2] išvedama nauja pagrindinių komponenčių metodo modifikacija.

2.5. Požymių išskyrimas pritaikant $HSIC_0$

Pritaikydami (2.6) įvertį autoriai [7] panaudodami tą pačią kaip ir pagrindinių komponenčių metodo išvedime naudojamą Lagranžo daugiklių maksimizavimo techniką sukūrė pagrindinių komponenčių analogą. Toliau pateikiamas metodo išvedimas.

Tarkime, kad $\mathcal{X} = \mathbb{R}^D$ yra stebėjimų erdvė, o Θ yra galimų klasių aibė. Klasės, kurioms priklauso stebėjimas $x \in \mathcal{X}$ yra Θ poaibis, kurį galime atvaizduoti kaip $|\Theta|$ matavimų dvejetainį vektorių y, kurio komponentė su reikšme 1 reiškia, kad stebėjimas priklauso tos komponentės nurodomai klasei, o komponentė su reikšme 0 reiškia, kad nepriklauso. Tuomet visos galimos klasės sudaro erdvę $\mathcal{Y} = \{0,1\}^{|\Theta|}$. Turėdami stebėjimus galinčius vienu metu priklausyti daugeliui klasių $S = \{(x_1, y_1), ..., (x_m, y_m)\}$, norime rasti funkciją $h : \mathcal{X} \to \mathcal{Y}$ galinčią prognozuoti klases naujiems stebėjimams su nežinomois y reikšmėmis.

Laikydami, kad stebėjimai ir jų klasės yra priklausomi, raskime mažesnio matavimo erdvę $\mathcal{F}$, kurioje priklausomybė tarp stebėjimų projekcijų šioje erdvėje ir stebėjimus atitinkančių klasių yra maksimali. Paprastumo dėlei laikykime, kad duomenis į $\mathcal{F}$ atvaizduojame tiesiniu atvaizdžiu $\phi(x) = P^T x$. Tuomet bandykime maksimizuoti priklausomybę tarp požymių $\phi(x) \in \mathcal{F}$ ir klasių reikšmių $y \in \mathcal{Y}$. Tam panaudokime empirinį HSIC įvertį (2.6).

Tuomet šiuo atveju skaičiuodami (2.6) parinkime branduolius taip, kad $k(x, x') = \langle \phi(x), \phi(x') \rangle = \langle P^T x, P^T x \rangle$, ir $l(y, y') = \langle y, y' \rangle$.

Kadangi (2.6) daugiklis priklausantis tik nuo stebėjimų skaičiaus maksimizavimui neturi įtakos, tai nagrinėkime tik reiškinį $\mathrm{tr}KHLH$. Pažymėkime $X = [x_1, ..., x_m]$ ir $Y = [y_1, ..., y_m]$. Tuomet $\phi(X) = P^T X$, $K = \langle \phi(X), \phi(X) \rangle = X^T P P^T X$, ir $L = Y^T Y$. Tuomet galime perrašyti maksimizavimo uždavinį kaip optimalios tiesinės projekcijos

$$P^* = \arg\max_P \operatorname{tr}(HX^T PP^T XHL)$$

radimo uždavinį.

Tarkime, kad norime sumažinti dimensijų skaičių iki $d < D$. Pažymėkime $P = [P_1, ..., P_d]$ – matricos P stulpelius-vektorius sudarančius naujos erdvės bazę. Kaip ir pagrindinių komponenčių metode, įtraukę ortonormalumo apribojimą gauname maksimizavimo uždavinį

$$\max_P \operatorname{tr}(HX^T PP^T XHL)$$

su apribojimu $P^T P = I$. Reiškinį pertvarkę turime

$$\operatorname{tr}(HX^T PP^T XHL) = \operatorname{tr}\left(\sum_{i=1}^d HX^T P_i P_i^T XHL\right) =$$
$$\sum_{i=1}^d \operatorname{tr}(HX^T P_i P_i^T XHL) = \sum_{i=1}^d P_i^T (XHLHX^T) P_i,$$

ir pritaikius Lagranžo daugiklių metodą išplaukia, kad optimalūs projekcijos vektoriai P_i^*, ($1 \leqslant i \leqslant d$) gaunami kaip matricos $XHLHX^T$ tikriniai vektoriai atitinkantys d didžiausių jos tikrinių reikšmių $\lambda_1, ..., \lambda_d$. Kadangi $XHLHX^T$ yra simetriška, tai visos tikrinės reikšmės yra realios. Norint išskirti požymius šiuo metodu, veiksmų skaičius yra $O(D^3)$. Optimalią projekcijų matricą P^* atitinkanti HSIC įverčio reikšmė yra $\sum_{i=1}^d \lambda_i$. Optimizavimo uždavinys susiveda į matricos $D \times D$ tikrinių reikšmių išvedimą, ir veiksmų aritmetinių veiksmų skaičius yra $O(D^3)$.

Viršuje esančioje analizėje autoriai kaip branduolio funkciją naudoja skaliarinę sandaugą. Jeigu toks paprastas tiesinis branduolys nepakankamas ryšiui tarp klasių aptikti, tai rekomenduoja naudoti sudėtingesnę branduolio funkciją. Kadangi L_{ij} atspindi ryšį tarp klasių y_i ir y_j, tai tokiu būdu išskirdami požymius randame pradinių duomenų atvaizdį $\phi : \mathcal{X} \to \mathcal{F}$ atspindintį jų ryšį su klasėmis.

2.6. Požymių išskyrimas pritaikant $HSIC_1$

Pritaikydami nepaslinktą HSIC įvertį (2.7) autoriai [2] panaudodami taip pat pagal Lagranžo daugiklių maksimizavimo techniką sukūrė pagrindinių komponenčių analogą. Toliau pateikiamas metodo išvedimas.

Tarkime, kad turime m stebėjimų $x_i \in \mathbb{R}^D$, o y_i yra priklausomi kintamieji. Pažymėkime $\phi(x) = P^T x$ – tiesinį atvaizdį į požymių erdvę, kur $P \in \mathbb{R}^{D\times d}$ ir $P^T P = I$. Stebėjimų nepriklausomiems kintamiesiems parinkime tiesinį branduolį $k(x, x') = \langle P^T x, P^T x\rangle = x^T P P^T x$, o priklausomiems kintamiesiems branduolį galime parinkti bet kokį. Tuomet atitinkamos branduolio matricos $K = X^T P P^T X$ ir $L = \{l(y_i, y_j)^m_{i,j=1}\}$. Pažymėkime matricas $A := \mathbf{1}\mathbf{1}^T$, $H := I - m^{-1}A$, $\tilde{K} := K - \mathrm{diag}(K)$, $\tilde{L} := L - \mathrm{diag}(L)$, bei $\Phi := \tilde{K}\tilde{L} + \frac{\tilde{K}A\tilde{L}A}{(m-1)(m-2)} - \frac{2\tilde{K}\tilde{L}A}{m-2}$. Ieškokime projekcijos P atvaizduojančios D-matavimų stebėjimus į $d < D$ matavimų požymius maksimizuodami HSIC įvertį (2.7). Kadangi

$$HSIC_1(X, Y) = \frac{\mathrm{tr}(\Phi)}{m(m-3)} = \frac{\mathrm{tr}(\Phi^T)}{m(m-3)} = \frac{\mathrm{tr}(\frac{\Phi+\Phi^T}{2})}{m(m-3)}, \tag{2.8}$$

tai norėdami rasti optimalią projekcijos matricą turime su apribojimu $P^T P = I$ maksimizuoti sekančią tikslo funkciją:

$$f(P) = \mathrm{tr}\left(\frac{\Phi + \Phi^T}{2}\right) = \mathrm{tr}((X^T P P^T X - D_K)\tilde{L} + \frac{(X^T P P^T X - D_K)A\tilde{L}A}{(m-1)(m-2)} -$$

$$\frac{(X^T P P^T X - D_K)\tilde{L}A + A\tilde{L}(X^T P P^T X - D_K)}{m-2}) =$$

$$\sum_{i=1}^{d} \mathrm{tr}(X^T P_i P_i^T X\tilde{L} + \frac{(X^T P_i P_i^T X)A\tilde{L}A}{(m-1)(m-2)} - \frac{X^T P_i P_i^T X\tilde{L}A + A\tilde{L}X^T P_i P_i^T X}{m-2}) -$$

$$\frac{\mathbf{1}^T\tilde{L}\mathbf{1}}{(m-1)(m-2)} \sum_{i=1}^{d} P_i^T X X^T P_i + \frac{2}{m-2}\sum_{i=1}^{d} P_i^T (\sum_{j=1}^{m} x_j x_j^T \sum_{k=1}^{m} \tilde{L}_{jk}) P_i =$$

$$\sum_{i=1}^{d} P_i^T X(\tilde{L} + \frac{A\tilde{L}A - \mathbf{1}\tilde{L}\mathbf{1}^T I}{(m-1)(m-2)} - \frac{\tilde{L}A - A\tilde{L} - 2diag(\tilde{L}A)}{m-2}) X^T P_i$$

čia $D_K = \mathrm{diag}(K)$. Maksimizuodami funkciją $f(P)$ Lagranžo daugiklių metodu gauname, kad optimalią P gautume kaip d tikrinių vektorių atitinkančių d didžiausių matricos

$$M = X(\tilde{L} + \frac{A\tilde{L}A - \mathbf{1}\tilde{L}\mathbf{1}^T I}{(m-1)(m-2)} - \frac{\tilde{L}A + A\tilde{L} - 2diag(\tilde{L}A)}{m-2})X^T, \tag{2.9}$$

tikrinių reikšmių.

Iš (2.8) matome, kad M yra simetriška, ir todėl visos jos tikrinės reikšmės yra realios. Optimalią matricą P atitinkanti $HSIC_1$ reikšmė lygi $\frac{1}{m(m-3)} \sum_{i=1}^{d} \lambda_i$, kur λ_i yra i-toji didžiausia matricos M tikrinė reikšmė. $D \times D$ matricos tikrinių reikšmių uždavinio veiksmų skaičius yra $O(D^3)$.

Taigi, kad kai matrica $\mathcal{M} = XX^T$, tai turime požymius pagal įprastinį pagrindinių komponenčių metodą, kai $\mathcal{M} = XHLHX^T$, tai turime požymius pagal paslinktą HSIC įvertį (2.6), o jeigu $\mathcal{M}$ yra (2.9) tai turime požymius pagal nepaslinktą HSIC įvertį (2.7).

3. TIRIAMOJI DALIS

3.1. $HSIC_0$ įverčio analizė dviejų klasių atveju

Panagrinėkime Hilberto-Šmidto kriterijaus $HSIC_0$ įverčiu paremtu požymių išskyrimo metodu gaunamą įvertį dviejų klasių atveju.

Metodo taikymo metu skaičiuojame matricą $\mathcal{M} = XHLHX^T$, kur $H = I - \frac{1}{m}\mathbf{1}\mathbf{1}^T$, $L = l(y_i, y_j)$, $i, j = 1, ..., m$, čia y_i, y_j - klasių reikšmės.

Nagrinėkime atvejį, kai branduolys l yra kanoninė skaliarinė sandauga: $l(y_1, y_2) = \langle y_1, y_2 \rangle$. Tuomet $L = YY^T$ yra branduolio matrica. Dviejų klasių žymimų $\{0, 1\}$ atveju, kai klasių vektorius-stulpelis apibrėžtas lygybe $Y = \{y_{ij}\}$, kur $i = \overline{1, n}$, $j = 1$ su sąlygom

$$y_{i1} = 1, \ kai \ i = \overline{1, n_1},$$

$$y_{i1} = 0, \ kai \ i = \overline{n_1 + 1, n},$$

čia n_1 – vieneto klasės elementų skaičius, n_0 – nulio klasės elementų skaičius ($n_1 + n_0 = n$), o $H = \{h_{ij}\}$, kur

$$h_{ii} = \tfrac{n-1}{n}$$

$$h_{ij} = -\tfrac{1}{n}, kai\ i \neq j$$

todėl $HY = \{\sum_{j=1}^{n} h_{ij} \cdot y_j\}$, kur

$$i = \overline{1, n}$$

$$y_j = 1,\ kai\ j \leqslant n_1$$

$$y_j = 0,\ kai\ j > n_1.$$

Taigi, $HY = \{hy_i\}_{i=1}^{n}$, kur $hy_i = \sum_{j=1}^{n} h_{ij} y_j = \sum_{j=1}^{n_1} h_{ij} 1 + \sum_{j=n_1+1}^{n} h_{ij} y_j =$

$$\tfrac{n_0}{n},\ kai\ i \leqslant n_1$$

$$-\tfrac{n_1}{n},\ kai\ i > n_1.$$

Kadangi H simetriška, ir $HLH^T = HYY^T H^T = (HY)(HY)^T$, tai $HLH = \{a_{ij}\}$, kur $a_{ij} =$

$$\left(\tfrac{n_0}{n}\right)^2,\ kai\ i \leqslant n_1,\ j \leqslant n_1$$

$$-\tfrac{n_0 n_1}{n^2},\ kai\ i \leqslant n_1,\ j > n_1$$

$$-\tfrac{n_1 n_0}{n^2},\ kai\ i > n_1,\ j \leqslant n_1$$

$$\left(\tfrac{n_1}{n}\right)^2,\ kai\ i > n_1,\ j > n_1$$

Kadangi $j = \overline{1, n}$ stebėjimų yra matricos X stulpeliai $\{x_j^{(i)},\ i = \overline{1, k}\}$, tai $XHY = \{xhy_{ij}\}$, kur

$$xhy_{i1} = \frac{n_0}{n} \sum_{j=1}^{n_1} x_j^{(i)} - \frac{n_1}{n} \sum_{j=n_1+1}^{n} x_j^{(i)} =$$

$$= \frac{n_0 n_1}{n} \overline{x}_{i1} - \frac{n_1 n_0}{n} \overline{x}_{i0} = \frac{n_0 n_1}{n} \left(\overline{x}_{i1} - \overline{x}_{i0}\right),\ i = \overline{1, k}, j = 1,$$

Čia $\overline{x}_{i1} = \frac{1}{n_1} \sum_{j=1}^{n_1} x_j^{(i)}$ ir $\overline{x}_{i0} = \frac{1}{n_0} \sum_{j=n_1+1}^{n} x_j^{(i)}$.

Kadangi $XHLHX^T = XHYY^TH^TX^T = XHY(HY)^TX^T = XHY(XHY)^T$, tai

$$\mathcal{M} = (XHY)(XHY)^T = \frac{n_0^2 n_1^2}{n^2} \left(\overline{x}_{11} - \overline{x}_{10}, ..., \overline{x}_{k1} - \overline{x}_{k0}\right)^T \cdot \left(\overline{x}_{11} - \overline{x}_{10}, ..., \overline{x}_{k1} - \overline{x}_{k0}\right)$$

Pažymėję $d_i = \overline{x}_{i1} - \overline{x}_{i0}$, $c = \frac{n_0^2 n_1^2}{n^2}$, $T = (d_1, d_2, ..., d_k)^T$ ir sudauginę matricas gauname, kad

$$\mathcal{M} = c \cdot TT^T = c \cdot \begin{pmatrix} d_1^2 & d_1 d_2 & \cdots & d_1 d_k \\ d_2 d_1 & d_2^2 & \cdots & d_2 d_k \\ \vdots & \vdots & \ddots & \vdots \\ d_k d_1 & d_k d_2 & \cdots & d_k^2 \end{pmatrix}$$

Imkime kitą branduolį, pavyzdžiui, $l = (y_i, y_j) = \exp(y_i - y_j)^2$. Tuomet $L = \{l_{ij}\}$, kur

$$l_{ij} = 1,\ kai\ y_i = y_j$$

$$l_{ij} = e,\ kai\ y_i \neq y_j.$$

Matricos H išraiška sutampa, todėl $HL = \{b_{ij}\}$, $b_{ij} =$

$$\frac{n_0}{n}(1-e),\ kai\ i \leqslant n_1,\ j \leqslant n_1$$

$$-\frac{n_0}{n}(1-e),\ kai\ i \leqslant n_1,\ j > n_1$$

$$-\frac{n_1}{n}(1-e),\ kai\ i > n_1,\ j \leqslant n_1$$

$$\frac{n_1}{n}(1-e),\ kai\ i > n_1,\ j > n_1$$

Taigi, $HLH = \{a'_{ij}\}$, $a'_{ij} =$

$$2\frac{n_0^2}{n^2}(1-e),\ kai\ i \leqslant n_1,\ j \leqslant n_1$$

$$-2\frac{n_1 n_0}{n^2}(1-e),\ kai\ i \leqslant n_1,\ j > n_1$$

$$-2\frac{n_0 n_1}{n^2}(1-e),\ kai\ i > n_1,\ j \leqslant n_1$$

$$2\frac{n_1^2}{n^2}(1-e),\ kai\ i > n_1,\ j > n_1$$

Išsikėlę $t = 2 - 2e$ matome, kad bet kurio iš šių branduolių atvejais matricos HLH išraiška skiriasi tik konstanta, ir gali būti užrašyta kaip $t \cdot TT^T$, kur T yra vektorius-stulpelis $\{\overline{x}_{i1} - \overline{x}_{i0}\},\ i = \overline{1,k}, j = 1$.

Nesunku pastebėti, kad taip bus bet kurio branduolio l su savybe

$$l(a,a) = l(b,b),\ kai\ a \neq b, \tag{3.1}$$

a, b - elementai žymintys stebėjimų klases atveju, ir tuomet konstanta $t = 2(l(1,1) - l(1,0))$. Kuomet $l(a,a) \neq l(b,b)$, bet $l(a,a) = 0$ arba $l(b,b) = 0$, konstanta $t =$

$$\begin{aligned} &l(a,a),\ kai\ l(b,b) = 0 \\ &l(b,b),\ kai\ l(a,a) = 0. \end{aligned} \tag{3.2}$$

Taigi, $HSIC_0$ paremto metodo matricos $\mathcal{M}$ įvertis dviejų klasių atveju yra stebėjimų komponenčių vidurkių skirtumų dekompozicija, padauginta iš konstantos. Konstanta priklauso nuo konkretaus branduolio.

Skaičiuojant matricą $\mathcal{M}$ be dekompozicijos, veiksmų skaičius $O(D+n^2)$, o su dekompozicija $O(D^2+n)$. Čia D - matavimų skaičius, n - stebėjimų skaičius.

3.2. $HSIC_0$ įverčiu gauta pirmoji pagrindinė komponentė dviejų klasių atveju

Sakykime $k = 2$. Tuomet matrica T^* lygi

$$T^* := TT^T = \begin{pmatrix} d_1^2 & d_1 d_2 \\ d_1 d_2 & d_2^2 \end{pmatrix}$$

Tikrinės reikšmės randoms iš lygybės

$$|T^* - \lambda I| = 0$$

$$\begin{vmatrix} d_1^2 - \lambda & d_1 d_2 \\ d_1 d_2 & d_2^2 - \lambda \end{vmatrix} = 0$$

$$(d_1^2 - \lambda)(d_2^2 - \lambda) - d_1^2 d_2^2 = 0$$

$$\lambda_1 = d_1^2 + d_2^2, \quad \lambda_2 = 0$$

Pirmoji pagrindinė komponentė $(v_1 \ v_2)^T$ yra tikrinis vektorius atitinkantis didžiausią tikrinę reikšmę $\lambda_1 = d_1^2 + d_2^2$.

$$(T^* - \lambda_1 I)(v_1 \ v_2)^T = 0, v_1^2 + v_2^2 = 1$$

$$\begin{pmatrix} -d_2^2 & d_1 d_2 \\ d_1 d_2 & -d_1^2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 0$$

Tuomet iš lygčių sistemos

$$v_1 = \tfrac{d_1}{d_2} v_2$$

$$v_1^2 + v_2^2 = 1$$

turime, kad

$$v_1 = \pm \frac{d_1}{\sqrt{d_1^2 + d_2^2}}, v_2 = \pm \frac{d_2}{\sqrt{d_1^2 + d_2^2}}$$

Tuomet koordinačių transformacijos, pirmoji komponentė lygi

$$z_1 = \frac{d_1}{\sqrt{d_1^2 + d_2^2}} X^{(1)} + \frac{d_2}{\sqrt{d_1^2 + d_2^2}} X^{(2)}.$$

4. PROJEKTINĖ DALIS

4.1. Uždavinys ir EKG duomenys

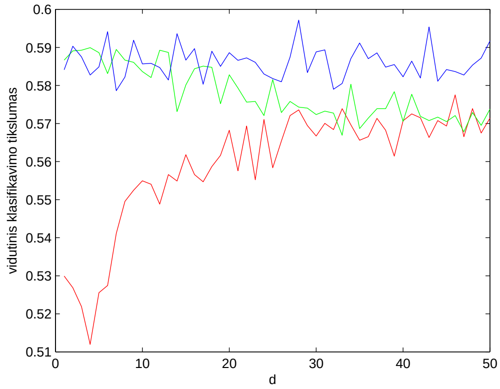
Iš aplinkotyros katedros ir gauti duomenys buvo sudaryti iš 305 stebėjimų, kurių kiekvienas yra sudarytas iš 183 komponenčių. 168 iš šių komponenčių sudaro nepriklausomus kintamuosius ir 15 iš kurių priklausomus kintamuosius. Visi nepriklausomi kintamieji yra kiekybiniai, o priklausomi kintamieji yra loginės reikšmės $\{1, 0\}$, apibūdinančios stebėjimo priklausomybę klasėms (stebėjimas gali tuo pačiu metu priklausyti vienai ar daugiau iš 15-os klasių).

Apibrėžiant funkciją $y_i := \bigvee_{i=1}^{15} y_j$ uždavinys suvedamas į vienos klasės binarinio klasifikavimo uždavinį.

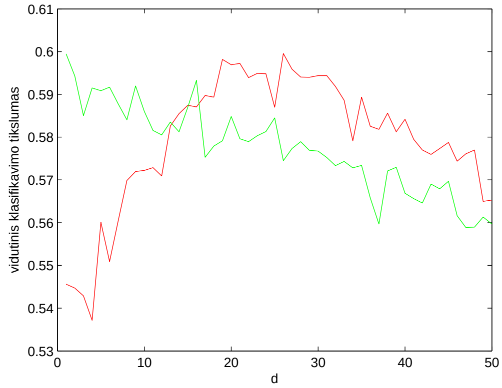
Iš pradžių palyginome gaunamus klasifikavimo rezultatus išskiriant požymius klasikiniu pagrindinių komponenčių metodu, ir $HSIC_0$ bei $HSIC_1$ paremtais požymių išskyrimo metodais. Tuomet palyginome skaičiavimo laiką reikalingą gauti $HSIC_0$ kriterijumi paremtą gauti matricai $\mathcal{M} = XHLHX^T$ tiesiogiai, ir naudojant dekompoziciją, bei kelis skirtingus branduolius.

4.2. Klasifikavimo rezultatų palyginimas

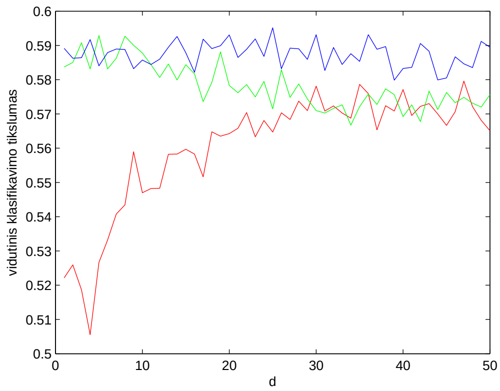
Klasifikavimo naudojant skirtingus požymių išskyrimo metodus vidutinį tikslumą empiriškai palyginome naudodami pacientų EKG duomenis gautus darbo praktikos metu. Su kiekvienu komponenčių skaičiumi d (abscisių ašys) kross-validavimo būdu atlikome po 100 bandymų suklasifikuoti 152 stebėjimus, ir įvertinome vidutinį klasifikavimo tiklsumą.



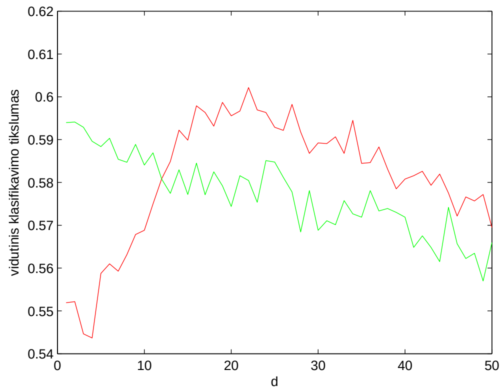
(a) kNN, $l(y_i, y_j) = \langle y_i, y_j \rangle$



(b) Logit, $l(y_i, y_j) = \langle y_i, y_j \rangle$



(c) kNN, $l(y_i, y_j) = e^{-0.1\sqrt{(y_i - y_j)^2}}$



(d) Logit, $l(y_i, y_j) = e^{-0.1\sqrt{(y_i - y_j)^2}}$

2 pav.: *Klasifikavimo su požymių išskyrimu tikslumo priklausomybė nuo išskirtų požymių naudojamų klasifikavimui skaičiaus. Raudona linija atitinka klasifikavimo tikslumą išskiriant požymius klasikiniu pagrindinių komponenčių metodu, žalia linija –* $HSIC_0$ *paremtu metodu, mėlyna linija –* $HSIC_1$ *paremtu metodu. Trumpinimas kNN reiškia k-artimiausių kaimynų metodą, o Logit – logistine regresija paremtą metodą. Šalia nurodyti požymių išskyrimui naudoti branduoliai.*

4.3. Skaičiavimo laiko palyginimas

Naudojome tuos pačius pacientų duomenis, pabandėme 5000 kartų rasti matricą $\mathcal{M}$ naudojant metodą su dekompozicija, ir be jos.

Bandymas	$l(y_i, y_j) = \langle y_i, y_j \rangle$	$l(y_i, y_j) = (\langle y_i, y_j \rangle + 1)^2$	$l(y_i, y_j) = e^{-0.1\sqrt{(y_i - y_j)^2}}$
Be dekompozicijos	208.7 sek.	214.1 sek.	381 sek.
Su dekompozicija	5.6 sek.	6.4 sek.	8.1 sek.

Naudodami branduolį $l(y_i, y_j) = e^{-0.1\sqrt{(y_i - y_j)^2}}$, ir fiksavę stebėjimų skaičių $n = 305$, pamėginome 500 kartų rasti matricą $\mathcal{M}$ dvigubindami matavimų skaičių D, pradedant $D = 84$.

D	Be dekompozicijos	padidėjimo kartai	Su dekompozicija	padidėjimo kartai
2688	762.7 sek.	2.94	199.7 sek.	3.76
1344	259.8 sek.	2.30	52.9 sek.	3.50
672	112.7 sek.	1.95	15.1 sek.	4.08
336	57.6 sek.	1.36	3.7 sek.	4.63
168	42.1 sek.	1.46	0.8 sek.	2.67
84	28.8 sek.		0.3 sek.	

Tuomet su tuo pačiu branduoliu, ir fiksuodami matavimų skaičių $D = 84$, pamėginome 500 kartų rasti matricą $\mathcal{M}$ dvigubindami stebėjimų skaičių n, pradedant $n = 305$.

n	Be dekompozicijos	padidėjimo kartai	Su dekompozicija	padidėjimo kartai
4880	7151.9 sek.	3.95	6.2 sek.	1.88
2440	1810.6 sek.	3.95	3.3 sek.	2.36
1220	457.4 sek.	3.94	1.4 sek.	2.33
610	116.2 sek.	3.95	0.6 sek.	2.0
305	29.4 sek.		0.3 sek.	

5. IŠVADOS IR REKOMENDACIJOS

5.1. Išvados

Darbe susipažinta su nauja ir plačiai pradedama naudoti duomenų analizės metodika – branduolių metodais.

Pritaikytos žinios apie pagrindinių komponenčių metodo principą, juo paremtiems naujiems požymių išskyrimo metodams įsisavinti. Metodai palyginti empiriškai.

Esant mažam klasifikavimui naudojamų komponenčių skaičiui $HSIC_0$ ir $HSIC_1$ įverčiais paremti metodai empiriškai buvo beveik lygiaverčiai, o esant didesniam naudojamų komponenčių skaičiui, $HSIC_1$ įverčiu paremtas metodas buvo pranašesnis. Klasikinis pagrindinių komponenčių metodas su mažu komponenčių skaičiumi duoda žymiai prastesnius rezultatus nei $HSIC_0$ bei $HSIC_1$ įverčiais paremti metodai.

Rasta Hilberto-Šmidto kriterijaus įverčiu $HSIC_0$ paremto metodo matricos dekompozicija. Dviejų klasių atveju $HSIC_0$ paremto metodo matricos $\mathcal{M}$ įvertis yra stebėjimų komponenčių vidurkių skirtumų dekompozicija, padauginta iš konstantos visiems branduoliams tenkinantiems savybę (3.1) arba (3.2) atvejį. Nuo branduolio priklauso tik konstanta dauginama iš matricos gautos skaičiuojant pagal dekompoziciją. Palyginti skaičiavimo našumai su dekompozicija ir be jos. Priklausomai nuo to, ar skaičiuojame su dekompozicija, ar be jos, skaičiavimo laikas žymiai skyrėsi: skaičiuojant be dekompozicijos skaičiavimo laikas proporcingas $O(D+n^2)$, o su dekompozicija $O(D^2+n)$, čia D - matavimų skaičius, n - stebėjimų skaičius.

5.2. Rekomendacijos

Norint gilintis į klasifikavimo uždavinį, rekomenduojame pasidomėti branduolių metodais, ir jų taikymais. Tam rekomenduojame [1] bei knygas [4], [5].

Klasifikuojant pagal EKG požymius, ir norint išskirti kuo daugiau informacijos į vieną kintamąjį, rekomenduojame naudoti $HSIC_0$ arba $HSIC_1$ įverčiais paremtus požymių išskyrimo metodus.

Koronarų stenozės prognozėje, pirmą $HSIC_0$ pagrindinę komponentę galima naudoti kaip informatyvų išvestinį rodiklį.

Kuomet matavimų skaičius labai didelis, tai rekomenduojame naudoti skaičiavimą be dekompozicijos, o kai stebėjimų skaičius labai didelis, ir branduolio funkcija tenkina savybes (3.1) ir (3.2), dviejų klasių atveju naudoti skaičiavimą su dekompozicija.

Šiame darbe skaičiavimo matricos dekompozicija rasta tik dviejų klasių atveju, tačiau šie metodai (dėl galimybės skaičiuoti branduolio reikšmes tarp vektorių) tinka ir daugiaklasiam požymių išskyrimui, todėl esant poreikiui rekomenduojame paieškoti šio metodo matricos dekompozicijos ir bendresniam atvejui.

Literatūra

1. Kernel Methods. http://onlineprediction.net/?n=Main.KernelMethods .

2. P. Daniušis, P. Vaitkus. Supervised Feature Extraction Using Hilbert-Schmidt Norms. *Vilniaus Universitetas*, 2008.

3. A. Gretton, O. Bousquet, A.J. Smola, B. Schoelkopf. Measuring Statistical Dependence with Hilbert-Schmidt Norms. *Max-Planck-Institute für biologische Kybernetik*, 2004.

4. B. Schölkpf, A.J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond.* The MIT Press, 2002. ISBN 0-262-19475-9.

5. J.S. Taylor. *Kernel Methods for Pattern Analysis.* Cambridge University Press, 2004. ISBN 0-521-81397-2.

6. Knyga: J. Venclovienė. *Statistiniai metodai medicinoje.* Vytauto Ddidžiojo universiteto leidykla, 2010.

7. Y. Zhang, Z. Zhou. Multi-label dimensionality reduction via dependence maximization. *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence*, 2008.

Literatūra

[1] Pagal Aronszajn. Kernel Methods. http://onlineprediction.net/?n=Main.KernelMethods, 2010.

[2] P. Daniušis and Vaitkus P. Supervised Feature Extraction Using Hilbert-Schmidt Norms. Vilniaus Universitetas, 2008.

[3] A. Gretton, O. Bousquet, A. Smola, and B. Schölkopf. Measuring Statistical Dependence with Hilbert-Schmidt Norms. *Max-Planck-Institute für biologische Kybernetik*, 2004.

[4] Bernhard Schoelkpf and Alexander J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. The MIT Press, 2002.

[5] John Shawe Taylor. *Kernel Methods for Pattern Analysis.* Cambridge University Press, 2004.

[6] J. Vencloviėnė. *Statistiniai metodai medicinoje.* Vytauto Ddidžiojo universiteto leidykla, 2010.

[7] Yin Zhang and Zhi-Hua Zhou. Multi-label dimensionality reduction via dependence maximization. *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence*, 2008.

literatura