„Statisztika"

bev.

most ≠ matek kurzus

≠ alkalmazott ipari/kutatói

inkább: ízelítő a többváltozós stat. és MI kapcsolatáról

⇓ Term.tud.

MI egyik része: - csoportosítás (clustering, pattern rec.)

- osztályozás (classification)
- modellezés (pl. regression)
- vegyes további (egy. vált. is)

klasszikus módszerek (átláthatóbbak, kevesebb adattal)

újabb módszerek (interpretáció? big data...)

deep learning egy ezek közül

előadás + tutoriálok + adatkiértékelés

magamról..., Szepesváry Pál...

tárgy átfedései

Hastie, Tibshirani, Friedman: The element of statistical learning

Cornell kurzus: Machine Learning for Intelligent Systems

Egyváltozós stat-valószínűségszámítás ⇐ kell a közös nyelvhez

kísérlet (random process) → esemény → valószínűségi változó

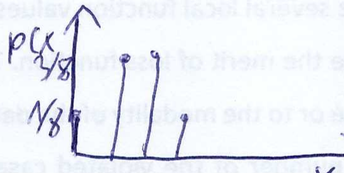
fv. értékkészlete

tudjuk ábrázolni

diszkrét v. folytonos ⇒ 0,1? −1,1? 1,2,3,4? 168,13cm összehasonlíts!

ξ, x

diszkrét val. vált. eloszlása



pl. 3 gyerek...

k valószínűsége

$p_k = \lim_{n\to\infty} \frac{n_k}{n}$ ← ez jön ki

axiómák vannak rá

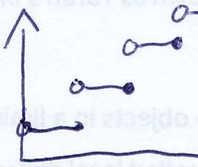
kumulatív:

$F(k) = p(x<k)$

$p(x \le k)$

melyik? $\left(= \sum_{x_i<k} p(x_i)\right)$

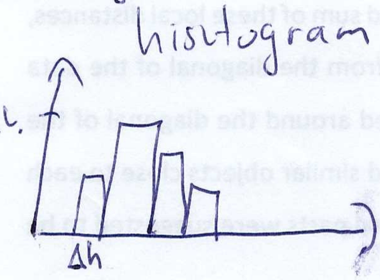
" eloszlás fv.



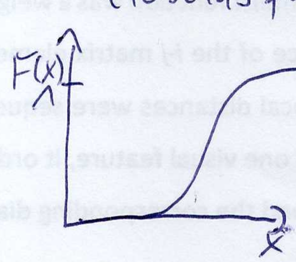
vagy ∘ • fordítva

folytonos:

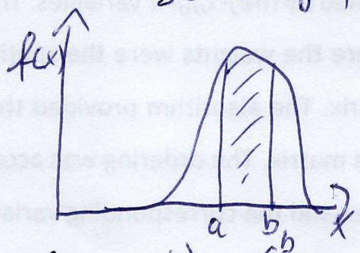
hisztogram



eloszlás fv



sűrűség fv



$p(a<x<b) = \int_a^b f(x)\,dx$

momentumai: számok, amik mondanak az eloszlásról valamit!

$$\int_{-\infty}^{\infty} \left(\frac{1}{\sigma}\right)^n (x-E(x))^n f(x)\,dx$$

normalizált

centrális

várható érték

$$E(x) = \int_{-\infty}^{\infty} x f(x)\,dx \qquad E(x) = \sum_{i=1} x_i p(x_i)$$

hol van

hoppa: $\bar{y} = \frac{\sum_{i=1}^{N} y_i}{N}$ ez mi? ← valszám ← statisztika

sokaság minta statisztika: mintából → sokaságról mondjuk valamit, hogy használjuk itt a valszámot

szórásnégyzet = variancia

$$\sigma^2 = \int_{-\infty}^{\infty} (x - E(x))^2 f(x)\,dx$$

terjedelem

$$s^2 = \frac{\sum_{i=1}^{N} (y_i - \bar{y})^2}{N-1} \qquad s = \sqrt{s^2}$$

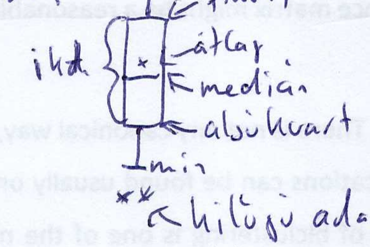
← ~~újfajta~~ újfajta példa a távolságokkal

$\bar{y}, s^2, s$ becsült a mintából, pl. becsült szórás

Robusztus becslés: hibás adat kevésbé terít önkre

medián, alsó-felső kvartilis, interkvartilis tár,

boxplot:



Tukey

← kilógó adat (med ± 1,5 ikt-on kívül)

← eldobott adat (± 2,5)

Eloszlások: binomiális, Poisson, exponenciális, normál, egyenletes...

χ^2, F, t származtatott

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-E)^2}{2\sigma^2}\right) \qquad F(x) = \text{erf fv.}$$

Ferdeség, csúcsosság, variáció koefficiens: $\frac{s}{\bar{y}}$

IV. levél.

Adattípusok:

metrikus (diszkrét, folytonos) - intervallum vagy arányos

kategórikus - nominális A, B, C - ← hot encodinggal binárissá

bináris 0,1

sorrendbeli +, ++, +++ → pl. vegyszerek, „zöldség"

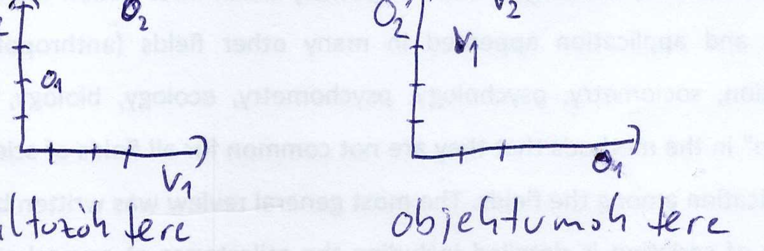
Többváltozós adatok tere:

X, y → $D^{N\times M}$ mátrix sor: objektum, eset

oszlop: tulajdonság (T, P, VH, cm^{-1})

vagy hipermátrix (tömb) 2D<

$$\begin{array}{c|cc} & v_1 & v_2 \\ \hline O_1 & 1 & 2 \\ O_2 & 3 & 4 \end{array}$$



Hiányzó adat - valamit beírni, ha kell

Centrálás - oszlopátlag 0 lesz

Skálázás - pl. min-max → 0,1, vagy osztás szórással

standardizálás → $\frac{y-\bar{y}}{s_y}$ → $x(-3,3)$, ha Gauss $\frac{d_{ij}-\bar{d}_j}{s_j}$

variancia $s_j^2 = \frac{\sum (d_{ij}-\bar{d}_j)^2}{N-1}$ $[0, +\infty)$

kovariancia $s_{jk} = \frac{\sum (d_{ij}-\bar{d}_j)(d_{ik}-\bar{d}_k)}{N-1}$ $(-\infty, +\infty)$

korreláció $r_{jk} = \frac{s_{jk}}{s_j \cdot s_k}$ $[-1, 1]$

r_{jk} = cos of v_1, v_2 vectors in object space

$r_{jk} = 1$ v. -1 ⇒ kollineáris

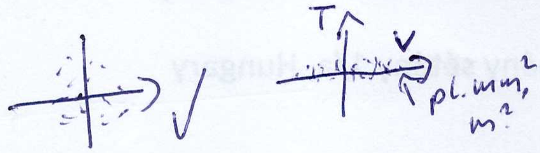
Gram mátrix: $G = D^T D$ kov. mátrix: $C^{M\times M} = \frac{D_c^T D_c}{N-1}$

korrelációs mátrix: $R^{M\times M} = \frac{D_{st}^T D_{st}}{N-1}$

Objektumok a változók terében:

távolság: $\underline{x}_{ij} = \underline{x}_i - \underline{x}_j$ ← sorok transzponáltja $\left(\boxed{\underline{x}_i^T} \right)$

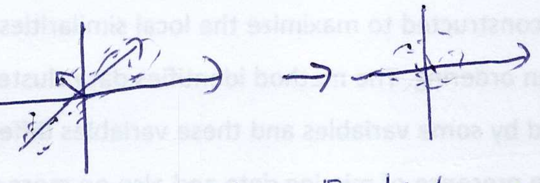
Euklideszi: $d_E = \sqrt{\underline{x}_{ij}^T \underline{x}_{ij}}$


pl. mm, m²

$d_{Pearson} = \sqrt{\underline{x}_{ij}^T S^{-1} \underline{x}_{ij}}$

$S = \begin{pmatrix} s_1^2 & & 0 \\ & \ddots & \\ 0 & & s_M^2 \end{pmatrix}$ varianciák

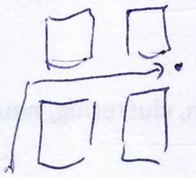
Mahalanobis-távolság és kerkorrelált változókra:


forgatással és skálázással!

Mah. táv. Eu. térben = Eu. táv. Mah. térben

$d_M = \sqrt{\underline{x}_{ij}^T C^{-1} \underline{x}_{ij}}$

Manhattan (city block táv.)



$d_{CB} = \sum_{n=1}^{M} |x_{in} - x_{jn}|$

távolság mátrix $(N \times N)$ $\begin{array}{c|c} & o_1 \quad \cdots \quad o_n \\ \hline \begin{matrix} o_1 \\ \vdots \\ o_n \end{matrix} & \begin{pmatrix} & & \\ & & \\ & & 0 \end{pmatrix} \end{array}$ $[0, +\infty)$

hasonlóság

$\frac{1}{1+d_{ij}}$ vagy $1 - \frac{d_{ij}}{d_{max}}$ $[0, 1]$

TV.terv 3.

bináris adatok távolsága és hasonlósága:

$$x_i = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad x_j = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} \Rightarrow \begin{array}{c|cc} & \text{1.obj } 1 & 0 \\ \hline 1 & a & b \\ 0 & c & d \\ \text{2.obj} & & \end{array} \rightarrow$$

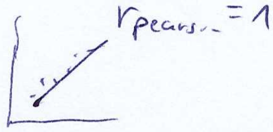
matching coeff.: $\frac{a+d}{a+b+c+d}$ ← kell-e ad?

Tanimoto koeff. $\frac{a}{a+b+c}$ dist: $\frac{b+c}{a+b+c}$

rang: sorrend $1 \dots N$ Spearman rangkorreláció két változó között

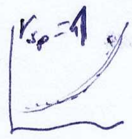
$\cancel{x_{ij}}$ $\underline{v_i} \rightarrow \underline{r_i}$ → r_{sp} = mint korreláció

$(x_{ij} \rightarrow r_{ij})$

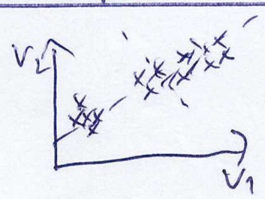


$r_{pearson} \approx$ linearitás

$r_{spearman} \approx$ nemlineáris, monotonitás



Főkomponens-elemzés:



? kell-e mindkét változó?
látom-e a két csoportot?

⇓

$\underline{v_1}, \underline{v_2}$ → más bázisra áttérni - - - → bázistranszformáció

Felbontás $D^{N \times M} = T \cdot P^T$

↗ score: új bázisban az értékek

loading ↖ régi változók súlya

$\underline{t}_i^T \underline{t}_j = 0$

$\underline{p}_i^T \underline{p}_j = 0$ ↖ merőleges

$$\underset{N}{\overset{M}{D}} = \underbrace{\underline{t}_1 \underline{p}_1^T + \underline{t}_2 \underline{p}_2^T + \dots}_{a} + E \quad \leftarrow \text{ha } a < M$$

↑ külső szorzat

↑ hibakomponens

matematikailag több megoldás, pl. SVD de

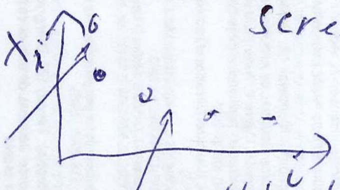
$C = \frac{D_c^T D_c}{N-1}$ diagonalizálásával (sajátérték egyenlet)

$C = \underbrace{Z \cdot \lambda}_{T} \cdot \underbrace{Z^{-1}}_{P^T}$ $\quad Z^{-1} = Z^T$ ha valós szimmetrikus (C az)

$a \le \text{rank}(C)$

Összvariancia: $\sum s_j^2 \rightarrow$ C: $\begin{pmatrix} s_j & & \\ & \ddots & \\ & & \ddots \end{pmatrix} \rightarrow$ átlós elemek összege megmarad

$\lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_M \end{pmatrix}$ ← összvariancia



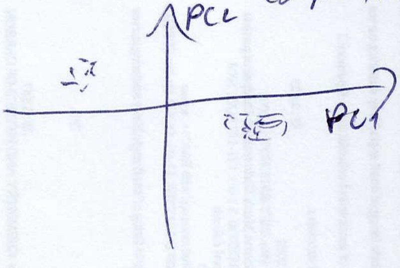
scree plot

vágj itt braking point

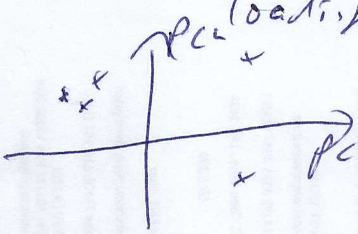
megmagyarázott összvariancia

$$\sum_{i=1}^{a} \lambda_i \Big/ \sum_{i=1}^{M} \lambda_i \rightarrow \text{pl. } 95\%$$

Mire jó: score plot obj. csoportosítás



loading változók



pl: pirolízis → 5 anyag → 2 főkomp → 2 mechanizmus 2 anyagcsop.

spektroszkópia, ~~[illegible]~~ összeldel vált. → PC-k a fő változások ⇐ modell rá

olajkutatás: két fő komp: felszíni, ~~[illegible]~~ tenger fenékközeli

gránit: fő kipárolgási folyamatok $(SiO_2 + K)$, $(Al_2O_3 + Na)$, $Fe_2O_3 - FeO$, $Na - K$

gránit, albit, oxid., alkáli balance

↓ térképre

piros Dacias

Clust 1.

PCA → változók számának csökkentése (dim. csökk.)

→ változók közti kapcsolatok

→ objektumok $\boxed{\text{csoportosítása}}$ (klaszterezés, mintázat felismerés)

↑ unsupervised learning (nem felügyelt, felügyelet nélküli; nem ellenőrzött...)

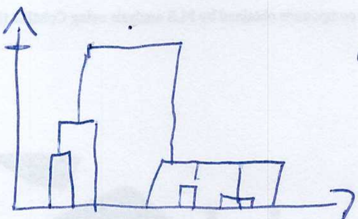
Hierarchikus klaszterezés:

- min t_{ij} a T távolságmátrixból (N×N a kezdetben)
- kit közeli klaszterbe vonása
- $(N-1)\times(N-1)$ T kiszámolása

Kérdések: klaszter-klaszter távolság? átlag, min, ~~max~~, ≃ linkage

távolság metrika (eu, Mah, man...)

$S \Rightarrow 1 - \frac{t_{ir}}{t_{max}}$



obs. sorrendje?

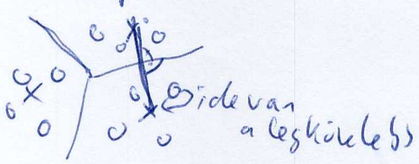
dendogram

K-mean klaszterezés:

$\underline{c}_g$ klaszter középek n_g adott oda tartozók

$$\min \sum_{g} \sum_{i \in g}^{n_g} (\underline{x}_i - \underline{c}_g)^T (\underline{x}_i - \underline{c}_g)$$ keresése

≃ Wigner-Seitz vagy Voronoi cella ⇒ tér felosztása



× $\underline{c}_g$

o $\underline{x}_i$

ki, hol, mennyire kompakt cellában, sziluett

hány csoport → nézegessük!

K-nearest-neighbour (Jarvis-Patrick) clust 2.

T → szomszédmátrix adott szomszédszámig (K)

i és j minimum M közös szomszédja (⇒ i és j egy csoport

pl. i, j egy csoport és j, l is ⇒ i, j, l egy csoport

pl.: $K = \frac{N}{2}$ $M = \frac{N}{4}$ ← hasraütésszerű

K, M próbálkozások

—

PCA: vizuális döntés síkon

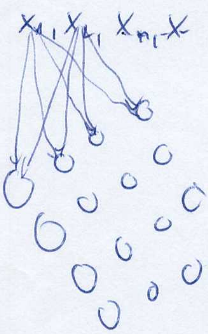
Hierarchikus, k-mean, knn: adott értékek táblázatos...

elmosódott határ: pl. fuzzy (élet...) klaszterezések → ki hová csak 90-os

—

legyen 2D leképezés: self-organizing map ∝ SOM

≈ nem felügyelt neurális háló – Kohonen finn úr



x_1, x_2, x_n, x, x_m bemenő többdimenziós adott $\underline{X}_1 \dots \underline{X}_n$ objektumra

– minden bemenő minden neuronnal összekötve egy súllyal ($\underline{w}_j$)

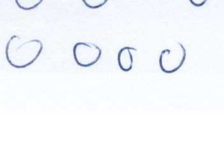
← síkban neuronok

súlyok változtatása úgy, hogy súlyértékek és adott $\underline{X}_i$ értékei hasonlóak legyenek: $\underline{X}_i$ a leghasonlóbb ~~súly~~ $\underline{w}_j$ súlyra hat + a környezet $\underline{w}_k$ is hatnak.

input vector (random choice)

min $(\underline{X}_i - \underline{w}_j)^T(\underline{X}_i - \underline{w}_j)$ j kiválasztása

Környék és kiválasztottakra: $w_k(s+1) = w_k(s) + \Theta_{táv}(t+s)(\underline{X}_i - \underline{w}_j)$



← adott objektumok 1-1 neuronra

↖ adott neuron súlyai vagy átlagos $\underline{X}$ bemenő adatai

Osztályozás - klasszifikáció

felügyelt tanulás

prediktor mátrix, feature deszkriptor: $X \to y$ vagy Y

y {-1,1}{0,1} bináris ← oneclass (spam)

{0,1,2,..5} osztályok, classes ← multiclass (Iris) (üveg, nize) (kőzet)

$y \in \mathbb{R}$ regresszió

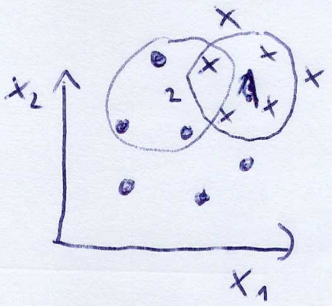
Célok: - ismert $X \to y$ -ra modellépítés, alkalmazás, ahol csak X ismert

- modell értelmezése: → súlyok, elválasztó felületek

→ adott objektumok ↔ modell kölcsönhatása

legegyszerűbb ötletek:

k-legközelebbi szomszéd:



pl. $k = 4 \to 1 \to X$

$2 \to 3$ o 1 x

honnan tudjuk, jó-e a modell?

osztályozási - klasszifikációs táblázat

pl: keresztellenőrzés: (pl. egyelem kihagyásos) CV-LOO, vagy teszt

valós ↘ / osztályozott →	1	2	3
1	$\boxed{50}$	0	0
2	0	$\boxed{48}$	2
3	0	1	$\boxed{48}$

□ jól

○ hibásan

példa $\Sigma = 150$

50/50/50
1 2 3

+ veszteségmátrix

$$\begin{bmatrix}0 & & ?\\ ? & 0 & \\ ? & & 0\end{bmatrix}$$

⇓

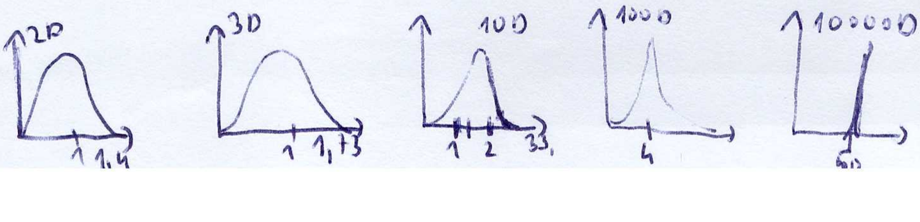
$\Rightarrow \Sigma \Rightarrow$ kockázat

hiba arány $\frac{\Sigma ○}{\text{összes}} \Rightarrow \frac{3}{150}$

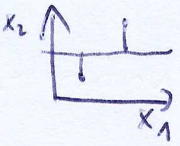
referencia hiba: $1 - \frac{N_{max}}{N}$ ← mindent oda

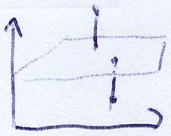
k-legközelebbi szomszédnál 2D, 3D, ... sok D ← távolság probléma

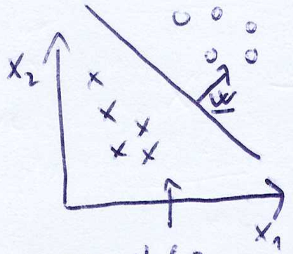
$d_{ij} = \sqrt{x_{i,j,1}^2 + x_{i,j,2}^2 + x_{i,j,3}^2 + \ldots}$ ← nagy távolság lesz, a "rokon".



pontok közti távolság helyett határoló síkokhoz, hipersíkokhoz való távolság:







$\underline{x}\underline{w} + b > 0$

$\underline{x}\underline{w} + b < 0$ ↑ mert nem megy át az origón

→ ha növelem +1-gyel a dimenziót.

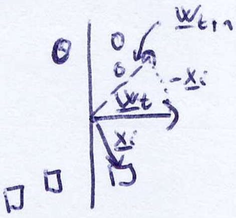
$\begin{bmatrix} \underline{x}_i \\ 1 \end{bmatrix} \leftarrow \underline{x}_i$ új

$\begin{bmatrix} \underline{w} \\ b \end{bmatrix} \leftarrow \underline{w}$ $\quad \underline{x}_i \underline{w} \lessgtr 0$

előjele adja!

Hogyan találjuk meg az elválasztó felületet?

Perceptron algoritmus: félreosztályozott elemek segítségével iteráljuk $\underline{w}$-t, amíg nem lesz félreosztályozott

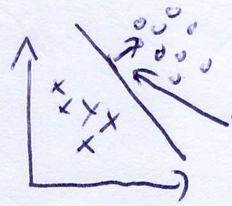


konvergál!

↓

max M lépésben

$M \leq \frac{1}{\gamma^2}$ ← széles margó a jobb



margó γ-táv a legközelebbitől

Támogató vektorok módszere Support vector machine

kb. ez, de maximális margó mindkét oldalra + további kényszerek....

↓ Kun- tömegalapján

↓ ~~krit.~~ SVM kritikus helyzetben levők alapján ← ezekre mondjuk a támogató vektor

nemilyen egyszerű, vannak akik rossz oldalon lesznek, sok további dolog....

Hol volt itt a valszám és a statisztika? Na most lesz!

Feltételes valószínűség (posterior) Bayes

$$P(A|B) = \frac{P(AB)}{P(B)}$$

$P(AB)$ ← szokták $P(A,B)$ ként is, pl. $P(X,y)$ közös

$P(B)$ ← prior, feltétel nélküli

$$P(A|B) = P(B|A)\cdot\frac{P(A)}{P(B)}$$ Bayes tétele

$P(X,y)$-t keressük a modellünkkel maximalni, modell paraméter $\underline{w}$

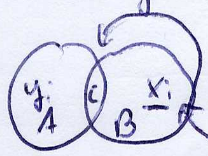
Maximum likelihood estimate (discriminative)

$P(\underline{y}|X,\underline{w})$ $\underline{w}$ keresése, ahol ez maximumot ad (közvetlenül)

$\underline{y} \to y_i$ $X \to \underline{x_i}$, függetlenség

$$= \prod_{i=1}^{n} p(y_i|\underline{x_i}, w)$$

Venn-diagram



$$P(y_i|x_i) = \frac{C}{B}$$

sok dimenziós, nem lesz ~~egy~~ sok közeli mintánk

← ezért mai irány

Maximum posterior estimate

$\underline{w}$ eloszlással rendelkező valószínűségi változó

$$P(\underline{w}|X,y) \approx p(\underline{y}|X,\underline{w})\cdot p(\underline{w}).$$

$$p(x_i|y_i) = \frac{C}{A}$$ ← alacsony dimenziós, sőt elméleti értékek pl. dobásokból $1/2$, $1/4$, $1/6$

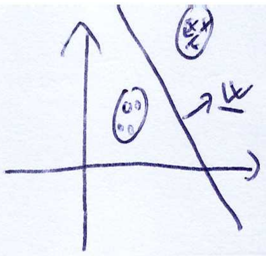
pontosabban

$$p(y_i|x_i) = \frac{p(x_i|y_i)\cdot p(y_i)}{p(x_i)}$$

$p(x_i|y_i)\cdot p(y_i)$ ← jobban becsülhető

$p(x_i)$ ← feltételenül

Mi volt a $\underline{w}$



$\left(\frac{\underline{w}}{n}\right)$ formában

$\underline{w}^T \underline{x}$ → előjele melyik csoport $y_i \{-1, +1\}$

MLE-vel: $p(y|\underline{x}_i) = \dfrac{1}{1+e^{-y(\underline{w}^T \underline{x}_i)}}$ discriminatív

⇓ mat min. keresés lesz log után (- előjellel)

$$\log\left(\prod_{i=1}^{n} p(y_i|\underline{x}_i, \underline{w})\right) = -\sum_{i=1}^{n} \log\left(1+e^{-y_i \underline{w}^T \underline{x}_i}\right)$$

$$\hat{\underline{w}}_{MLE} = \arg\min_{\underline{w}} \sum_{i=1}^{N} \log\left(1+e^{-y_i \underline{w}^T \underline{x}_i}\right)$$

logisztikus regresszió

MPE-vel

$$\hat{\underline{w}}_{MAP} = \arg\min_{\underline{w}} \sum_{i=1}^{n} \log\left(1+e^{-y_i \underline{w}^T \underline{x}_i}\right) + \underset{\overset{\uparrow}{\frac{1}{2\sigma^2}}}{\lambda} \underline{w}^T \underline{w}$$

generatív

$\underline{w} \sim \mathcal{N}(0, \sigma^2 I)$

lényeg: + tag

+ kismintánál fontos

+ segít feltételezéssel eloszlásra

Naive Bayes

Diszkriminancia analízis - régi módszer (régi alapmódszer)

felt. adatok csoporton belül Gauss eloszlásúak (több dim.)

g osztály $\underline{c}_g$ ← centruma C_g kovarianciája

felt. valószínűség maximálása, hogy odatartozik:

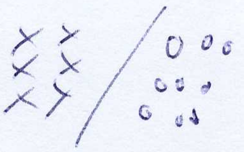
$$p(g|\underline{x}_i) = \max_g \frac{n_g/N}{|C_g|^{1/2}\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}(\underline{x}_i - \underline{c}_g)^T C_g^{-1} (\underline{x}_i - \underline{c}_g)\right)$$

2^2 Mahalanobis táv.

log. formában szokás keresni

diszkriminációs távolság

i-dik elem megfelelő csoporthoz tart.



határnál $D_A \approx D_B$

C_g egyforma Lineáris diszkriminancia analízis

LDA

C_g csop.-ként különböző Kvadratikus diszk. anal.

QDA

Döntési fák módszere — CART-classification

- Fokozatos osztogatása az objektumoknak egy-egy változóbeli érték alapján. Osztás alapja: legyen egy nagy és egy kicsi diverzitású csoport

nagy: heterogén az osztályokra

kicsi: homogén az osztályokra

diverzitás mértéke: Gini-index

$n_1, n_2, n_3 \dots n_m$ osztályelemek száma $\sum_{i=1}^{m} n_i = n$ $\quad p_i = \frac{n_i}{n}$

$$Gini = \sum_{j=1}^{m} \sum_{k \neq j}^{m} p_j p_k = \frac{1}{n^2} \sum \sum n_j n_k$$

$Gini = 0$ ← csak egy osztály van az adott vágás után az egyikben

$$\max Gini = \frac{\left(1 - \frac{1}{m}\right)}{2} \rightarrow Gini / Gini_{max} \in [0,1]$$

$Gini \Leftrightarrow$ Shannon $\quad H = -\sum_{i=1} p_i \ln p_i$

Gini: lineáris súly

Shannon: logaritmikus súly

$p_1 = 1 \quad p_{i \neq 1} = 0$

$\Downarrow$

$H = 0$

$p_i = \frac{1}{m} \quad H = \max$