Felügyelt tanulás 2. – regresszió ≈ modellezés, ha

$X \rightarrow y$ (vagy Y) $y_i \in \mathbb{R}$

célok: – modell alkalmazása, ahol csak X ismert
– modell alkalmazása, ahol csak y ismert (kalibráció)
– modell értelmezése

lineáris, nemlineáris → adott fv.
↘ nehezebben átlátható pl. neurális háló

Sokszor próbálgatás, de hivatalosan:

1) X, y azonosítása; függvénykapcsolat hipotézis: ma ~~ma~~ termtud. hálter vagy fekete doboz?; hiba tulajdonságai:
ε_i homo- vagy heteroszkedasztikus, ismert-e, $cov(\varepsilon_i, \varepsilon_j)$; hiba X és/vagy y;
becslési kritérium pl. $\sum |\hat{y}_i - y_i|$, $\sum (\hat{y}_i - y_i)^2$ – MLE, MPE
extrapoláció-interpoláció
2) paraméterezés
3) modell validálása

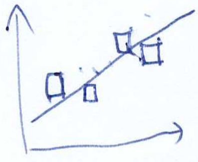
Legismertebb eset: egyenesillesztés – lineáris regresszió $X \rightarrow y$
least squares regression

X determinisztikus (hibamentes)
hiba csak y-ban
$E(\varepsilon_i) = 0$, ε_i ismert, $cov(\varepsilon_i, \varepsilon_j) = 0$ $i \neq j$-re

$\underline{x}_i - \underline{y}_i$ N is mert pár

cél fv: $Q = \sum_{i=1}^{N} (y_i - \underbrace{f(x_i, p)}_{\hat{y}_i \leftarrow \text{modell}})^2 \cdot w_i$ $\quad w_i = 1/\sigma_i^2$ ← súly

ábra



min

$f(x)$: $y = ax + b$

↓ súlyozatlanra

$Q(a,b) = \sum_{i=1}^{N} (y_i - ax_i - b)^2 \rightarrow \frac{\partial Q(a,b)}{\partial a} = 0$ parc. der. = 0

$\frac{\partial Q(a,b)}{\partial b} = 0$ normál egyenletek

$\bar{x}, \bar{y}$ átlagok

$$a = \frac{\sum (x_i y_i - \bar{x}\bar{y})}{\sum (x_i x_i - \bar{x}\bar{x})} \qquad b = \bar{y} - a\bar{x}$$

Miért négyzet? MLE-becslés norm. eloszlás hibával, σ_y^2 szórással

val. sűrűség fv. → $\rho_i(X|y_i) = \frac{1}{\sqrt{2\pi}\,\sigma_y} \cdot \exp\left(-\frac{(y_i - f(x_i,p))^2}{2\sigma_y^2}\right)$

$L = \prod_{i=1}^{N} \rho_i(x_i|y_i) = \ldots$ ← ezt maximálni akarjuk

maximális → $L' = \ln L = \ldots - \frac{1}{2} \underbrace{\sum \left(\frac{(y_i - f(x_i,p))^2}{\sigma_y^2} \right)}_{\text{minimalizálni}}$

ezekkel p torzítatlan $E(p)$ jó

konzisztens $N \to \infty$ KV

hatásos (minimális σ_p-t kapunk)

Többváltozós eset MLR

$y = p_0 + p_1 x_1 + p_2 x_2 + p_3 x_3 + \dots + p_m x_m = \underline{X}^T \cdot \underline{p}$

vagy $\hat{\underline{y}} = X \underline{p}$

$$X = \begin{pmatrix} 1 & x_{1m} \\ 1 & \\ \vdots & \\ 1 & x_{nm} \end{pmatrix}$$

konstanshez tartozik

$N \geq M+1$

$N = M+1$ egyértelmű megoldás, lineáris egyenletrendszer $\det(X) \neq 0$

$\exists y_i \neq 0$

Valóságban $\underline{y} \tilde{=} X \underline{p}$ hiba miatt

$$Q = \sum_{i=1}^{N} (\hat{y}_i - y_i)^2 = (\hat{\underline{y}} - \underline{y})^T (\hat{\underline{y}} - \underline{y}) \Rightarrow$$ sík, hipersík illesztése, lehűlési hűtés

memoritális levezetés: $X \underline{p} \tilde{=} \underline{y}$

$$X^T X \underline{p} = X^T \underline{y}$$

$$\underbrace{(X^T X)^{-1} (X^T X)}_{E} \underline{p} = (X^T X)^{-1} X^T \underline{y}$$

súlyozott: $\underline{p} = (X^T W X)^{-1} X^T W \underline{y}$ $\quad W = \begin{pmatrix} w_1 & & 0 \\ & \ddots & \\ 0 & & w_n \end{pmatrix}$ $\quad w_i = \frac{1}{\sigma_i^2}$

Kalibráció: cél modell felhasználása $\underline{x}$-hez tartozó $\underline{y}$ ismert

$\underline{y}$ helyett Y (több változós, pl. spektrumcsúcsok, pl. spektrumcsúcsok)

X komponens koncentrációk, S kalibrációs mátrix (reg. együtthatók)

modell: $Y = XS + E$

1.) S helyett $\hat{S} = (X^T X)^{-1} X^T \cdot Y$ kalib.

2.) $\hat{\underline{x}} = (\hat{S} \hat{S}^T)^{-1} \hat{S} \cdot \underline{y}_{mért}$ alkalmazás

$\hat{\underline{x}}$: keresett koncentráció

Miért nem megy sokszor az MLR? Kollinearitás

$\beta = (X^T X)^{-1} X^T y$ ~~korrelációs~~ kovariancia mátrix: $C = \frac{1}{N-1} X_c^T X_c$

← két hasonló sor

kvázi szinguláris ~~det(C)~~ $\det(X^T X^{-1}) \approx 0$ $R = \frac{1}{N-1} X_{st}^T X^T$

r

↗ két változó vektorainak cosinusza az objektumok terében

Hogyan kezeljük:

Változó szelekció:
- hasonlókból csak egyet hagyunk meg
- átlagolások...

Ridge regresszió $\beta \neq (X^T X)^{-1} X^T y$ helyett $\beta = (X^T X + \lambda E)^{-1} X^T y$

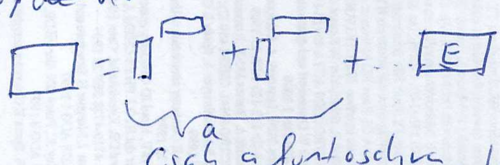
λ függvényében mit látunk (keresztell.)

MPE-nél láttuk: $P(\beta)$-re eloszlást feltételez, $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\beta^T \beta}{2\sigma^2}}$ alakút

λE-t általánosan a fizikusok regularizációnak nevezik, pl. Tikhonov reg: $(A^T A + \Gamma^T \Gamma)^{-1}$

↑ Tikh λE

Főkomponens regresszió; ~~de ko~~ (PCR)



csak a fontosakra, kémia: visszamérés kell!

PLS - parciális legkisebb négyzetek módszere

PCR-ben főkomponensek csak X alapján

PLS X és y vagy X és Y alapján
(PLS1) (PLS; PLS2)

látens változók (≈ főkomponensek) X és Y térben is

- maximális kovarianciájú pár kikeresése X - Y -ok közül
- illesztés erre, maradék számolása

eredmény ugyanúgy lineáris, legfőbbet hasonált lineáris regresszió, pl. sok spektrum érték közül csak a hulló

$$X = T \cdot P^T + E \quad \leftrightarrow \quad Y = UQ^T + F$$

(E: hib. elhagyolás; F: hiba, elhagyolás)

score loading

↑ jól korrelálló párok keresése...

MLP további variációi:

polinomiális $(1 \; x \; x^2 \; x^3)$

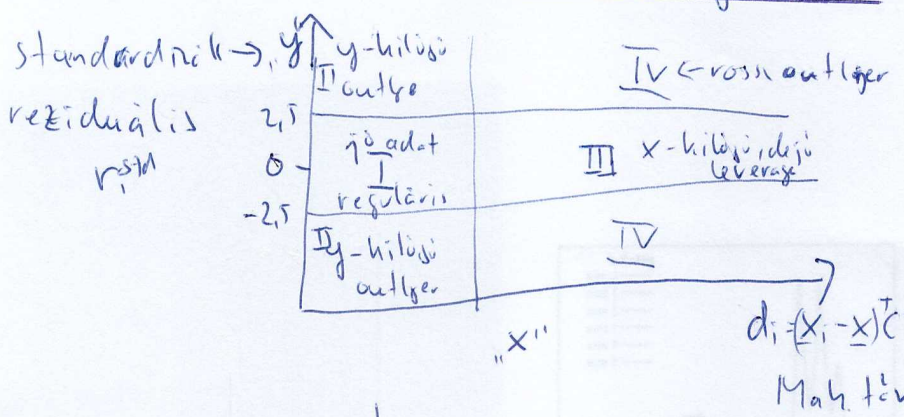
függvénnyel $y = c_0 + c_1 f_1(\underline{x}) + c_2 f_2(\underline{x}) + \ldots$

általánosított regr.: Mahalanobis-térben

GLS $\underline{b} = (X^T C^{-1} X) X^T C^{-1} y$

↑ súly helyett

Hibás adatokra – robusztus regresszió



↓

- úgy dolgozni, hogy csak I és III maradjon meg
- pár kilógó adat eldobása.
- más becslő fv. (nézzeke ki nagy hibánál)

X-y-ban is pontos/hibás 1 vált: (Total, Deming, ...)

$X_i \to \sigma_{x_i}^2 \qquad y_i \to \sigma_{y_i}^2 \qquad y = ax + b$

$$Q(a,b) = \sum_{i=1}^{N} (y_i - ax_i - b)^2 \cdot w_i = \sum_{i=1}^{N} \frac{(y_i - ax_i - b)^2}{\sigma_{y_i}^2 + a^2 \sigma_{x_i}^2}$$

hogyan

$$\widehat{var}(y_i - ax_i - b) = \widehat{var}(y_i) + a^2 \widehat{var}(x_i)$$

↑

igazából ez a jövő, de még csak megtört a leírásokban

Nemlineáris regresszió ismert függvényalakkal:

≈paraméterek becslése

$f(X,\beta)$ $\quad Q=(\underline{y}-\underline{f}(X,\beta))^T W(\underline{y}-\underline{f}(X,\beta))$

$$\begin{pmatrix} f_1(\underline{x}_1,\beta) \\ f_2(\underline{x}_2,\beta) \end{pmatrix}$$

megoldás J-vel Jacobi mátrix

$$J=\begin{pmatrix} \frac{\partial f(\underline{x}_1,\beta)}{\partial p_1} & \cdots & \frac{\partial f(\underline{x}_1,\beta)}{\partial p_m} \\ \vdots & & \\ \frac{\partial f(\underline{x}_n,\beta)}{\partial p_1} & & \frac{\partial f(\underline{x}_n,\beta)}{\partial p_m} \end{pmatrix}$$

Gauss-Newton $\quad \underline{\beta}^{k+1}=\underline{\beta}^k+(J^TWJ)^{-1}J^TW(\underline{y}-\underline{f}(X,\beta^k))$

vagy

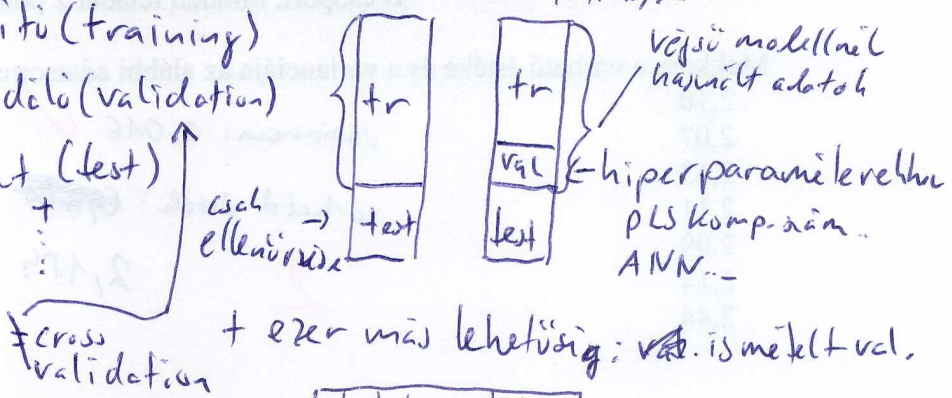
-Marquardt $\quad (J^TWJ)^{-1} \to (J^TWJ+\lambda E)^{-1}$

iterativ, mert nem lineáris, konvergencia nem biztos

Modellezés - elemi követelmény: modell validálása

adatok felosztása: tanító (training)
validáló (validation)
teszt (test)

(OECD guidance on QSAR) de nincs általános séma



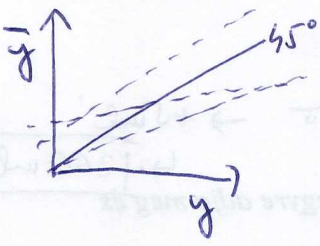
+ ezer más lehetőség: ~~val~~. ismételt val.

validációs paraméterek számításának alapja négyzetösszegek

$$TSS = \sum_{i=1}^{N_{tr}} (y_i - \bar{y})^2 \qquad RSS = \sum_{i=1}^{N_{tr}} (y_i - \hat{y}_i)^2 \qquad MSS = \sum (\hat{y}_i - \bar{\hat{y}})^2$$

~~PRESS~~ „teljes" lineáris regressziónál $TSS = MSS + RSS$, $\bar{\hat{y}} = \bar{y}$ másutt nem! ← ez okoz kavarodást, mert a rengeteg mutató!

Pearson r - korrelációs e.h. nemigazán jó, csak MLR ~~r~~



$r = 1$ mindre
$r = -1$
} r^2 véghipp. de MLR-re jó, mert $\bar{y} = \bar{\hat{y}}$...

átlagos négyzetes eltérés gyöke **RMSE** (root mean square error) $= \sqrt{\frac{RSS}{N}}$ $\Leftarrow$ $s = \sqrt{\frac{RSS}{N-1}}$ $0 \le RMSE$

determinációs együttható coefficient of determination. $R^2 = 1 - \frac{RSS}{TSS}$ $R^2 \le 1$

$R^2 = 0$ = átlag ugyanolyan jó!

(MLR-nél: $0 \le R^2 \le 1$ csak itt
$= r^2$
$= \frac{MSS}{TSS}$ ← modelláltal megmagyarázott)

Mindezekhez lehet mindegyik halmazra

„belső" validáció (modell paramétereinek-hiperparamétereinek meghatározásához használt adatokon)

- illeszkedés jósága (goodness of fit)

  $RMSE$, R^2 ... (F at MLR mél... konkordancia ...)

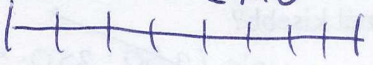
- robusztusság: a használt adatokra mennyire értelmes keresztellenőrzéssel (vagy bootstrappel)

$$PRESS = \sum (y_i - y_{(-i)})^2$$

↑ö nem vesz részt a modellben épp akkor

$$Q^2 = R^2_{val} = 1 - \frac{PRESS}{TSS} \qquad RMSE_{val} = \sqrt{\frac{PRESS}{n}}$$

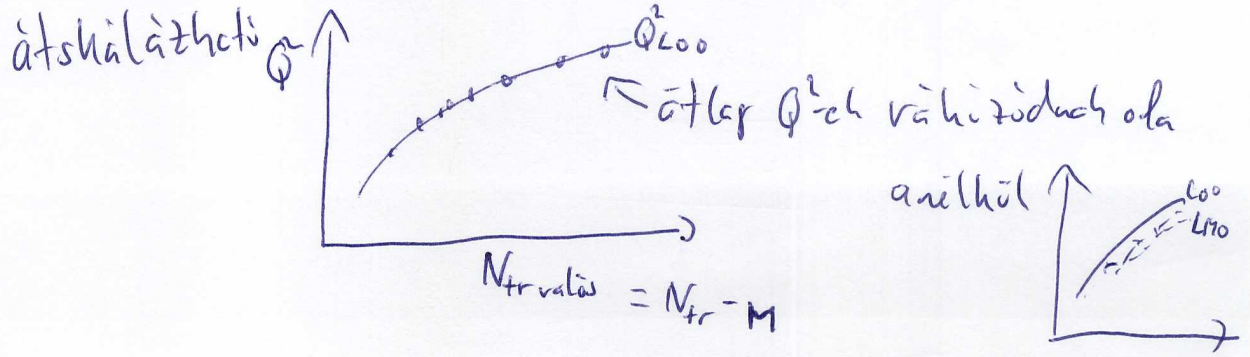
Leave-one-out CV
LOO
minden egyes mérés:
↓
N modell építése

Leave-many-out CV
LMO
LMO5%, LMO10%...

pl. 10 rész
90%-ra modell → 10% használ,
minden egyes mérés
N/10 modell építése

$-\infty < Q^2 \leq 1$

MLR-nél $-\infty < Q^2 \leq R^2 \leq 1$

van-e különbség Q^2_{LOO} és Q^2_{LMO} között? Nincs



„Külső" validáció – predikció, jóslás

teszt halmaz – sehol se használjuk a modell építésénél

$$RMSE_{test}\ (RMSE_{pred}) = \sqrt{\frac{\sum_{test}(y_i - \hat{y}_i)}{N_{TS}}}$$

$$R^2_{test} = Q^2_{F2} = 1 - \frac{RSS_{test}}{TSS_{test}}$$

Rengeteg más mutató van: – pl. négyzetösszeg helyett

varianciával: pl. (TSS/n-1) → adjusted R^2-ek, Q^2_{F1}, Q^2_{F3}

sokszor nem szemléletesek, de pl. ezzel jobban látható, hogy egyszerűbb vagy bonyolultabb modellt használjunk

– absz. értékesek ~~( )²~~ → $|y_i - \hat{y}_i|$

– zavarosabbak → néha félreértése... concordance corr. coefficient (Lin)

színesek... RO_{pi}

OECD-t én szeretem.

– adott helyen fontos: pl. szabványban: – bármelyik pontnál max ±15% eltérés

– 0.95 R^2_{tr} 0.99 R^2_{ts} ~~kalib~~

– visszanyerés (visszaszámolás anyagmennyiségre)

pl. ±5%

VAL-4

Bias - variance tradeoff ≈ kompromisszum

Levezethető egyszerű módon az alábontás

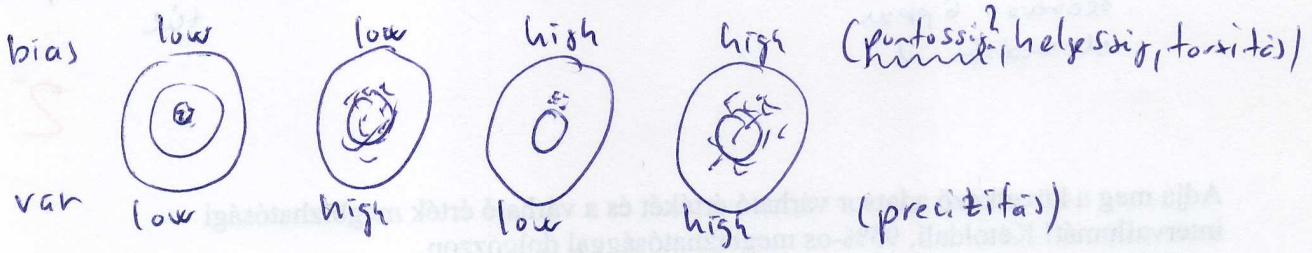
$E(\text{tesztkészlet hibája}) = E(\text{variancia}) + E(\text{zaj (noise)}) + E(\text{torzítás})^2$

Variancia forrása: mi van, ha más tréning halmazon dolgozunk.
Mennyire specifikus a modell (illeszkedés) erre a halmazra? ← paraméterekhez

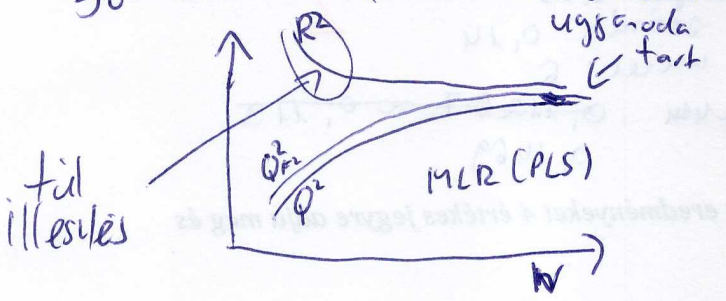
Bias: Mennyire tökéletlen a modell, mert hiányos, még végtelen- nagy tréningre is? ← modellalkotáshoz

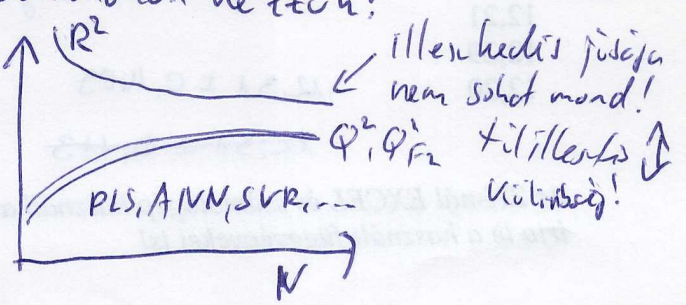
Zaj-hiba: Adatok belső zaja és hibája. ~~Hiba~~ Prediktorok tökéletlenségi-sűrűség, rossz mérhetőség. ← mérésekhez előtt

Analitikai kémiai analógia:



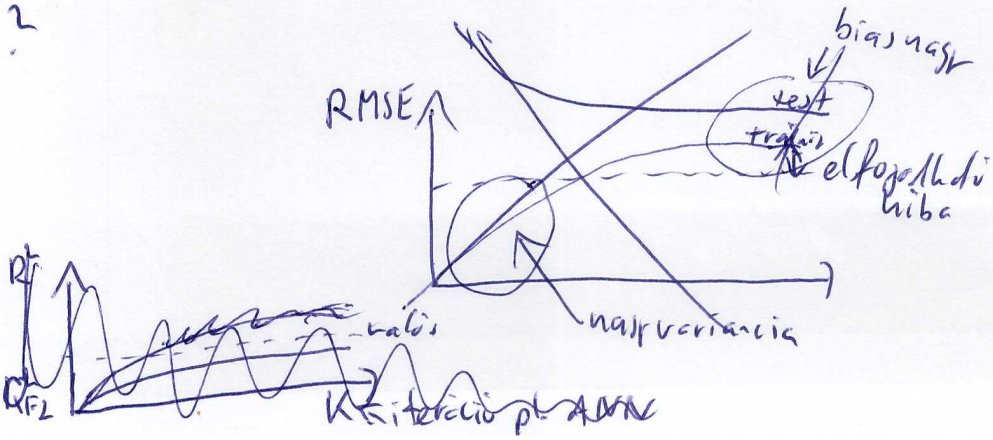
Jól követhető, ha különböző (tanító) mintaszámokon nézzük:





Mi, hogyan javítható?

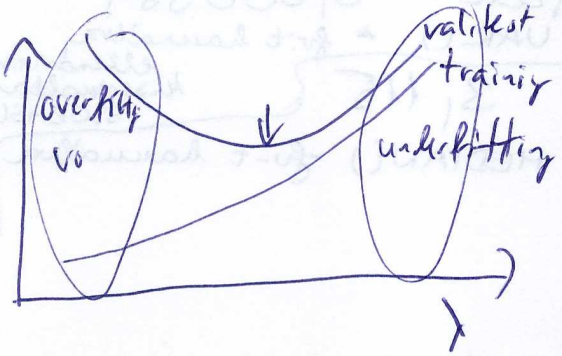
1) Variancia
- nagy mintaszám
- early stop



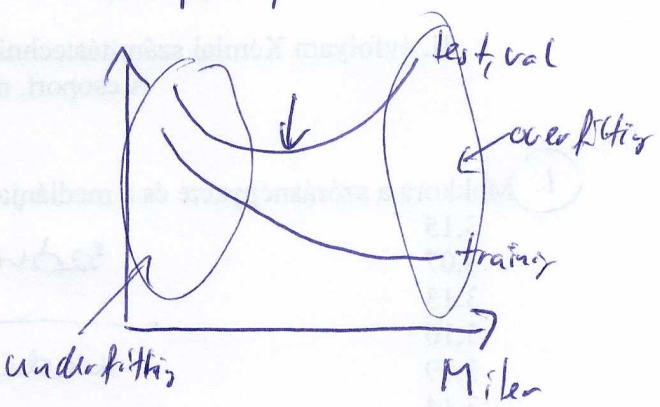
- regularizáció

pl. Ridge, MPE

$$\underline{w} = (X^T X + \lambda E)^{-1} X^T y$$



early stop:



- bagging-konszenzus modell

több tanító halmaz variánsrajlesztési →

az eredmény ezek összege

CART → random forest

módszer: bootstrap

Münchausen báró...

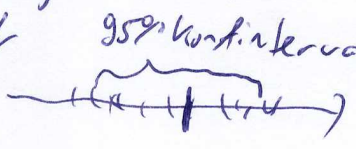
visszatevéses húzással $N \rightarrow N$ elemű új mintákat készítünk

T_f: ≈ populáció (pedig minta) ↑ minta

- pl. átlagok számolása belőle:

nem az összes, csak mondjuk 100-1000

átlagok eloszlása



- itt $N_{tr} \rightarrow$ 100 db új N_{tr} ismétléssel, 100 modell keverése

↓

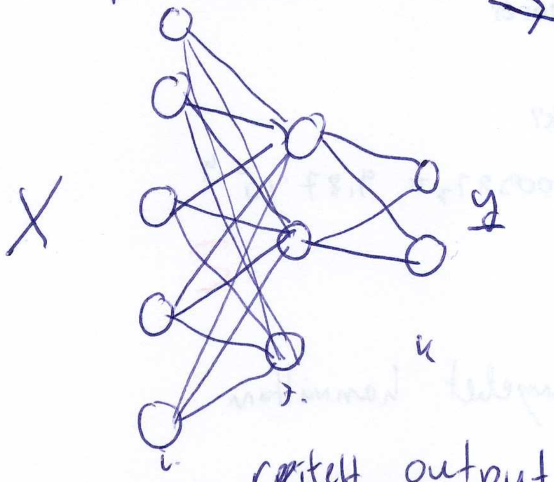
konszenzus, bagging

- mutatók: out of bag $Q^2 \approx$ robusztusságra

Mesterséges ideghálók – neurális hálózatok.

Alap modell:

adat → w_{ij} súly vagy w_{jk}



input rejtett réteg output

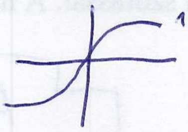
rejtett rétegbeli neuronjában (j-dik)

$$\to \sum w_{ij} x_{n,i} + b \to f(\;) \to$$

↑ n. bemenő i-dik változója

f típusai (ha $f(u)=u$ lineáris)

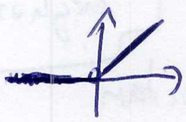
$$\tanh(u) = \frac{e^{u}-e^{-u}}{e^{u}+e^{-u}}$$



tangens hiperbolikus

$$\text{sigmoid}(u) = \frac{1}{1+e^{-u}}$$

relu: $u<0$: 0; $u=0$: N/A; $u>0$: u



elvileg nem szakadásos fv. leírható egy réteggel, de sok ∃!

használat:

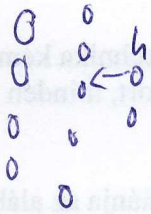
- ismert X, y párok ⇒ rengeteg súly optimálása, hogy $\hat{y} \approx y$, $\sum(\hat{y}-y)^2$ minimuma
- új X-ekre y-ok kiszámolása

  $y \in \mathbb{N}$ osztályozás

  $y \in \mathbb{R}$ regresszió

tanulás:

backpropagation alg.



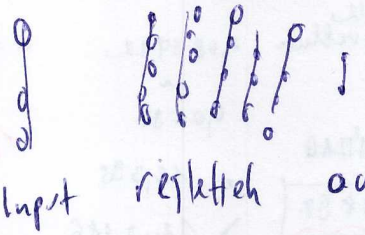
rengeteg súly, márpedig is $1 \times 3 + 3 + 3 \times 4 (+4) = N_w$

N fügyetlen adat, vitkán igaz hogy $N > 0{,}1 N_w$

backpropagation → alapvetően lokális minimumokat találó meredekség + sztochasztikus megoldások (SGD, ADAM)
+ early stop + bagging } min. var.

Mélytanulás, deeplearning

sok réteg:

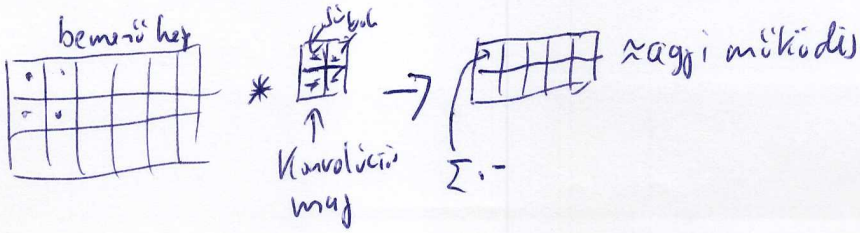


nagyon összetett dolgokat tud leírni, de rengeteg optimálandó súly

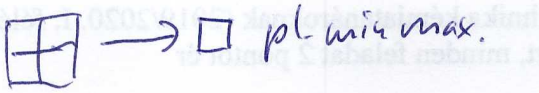
Konvolúciós neurális hálók képfeldolgozásra:

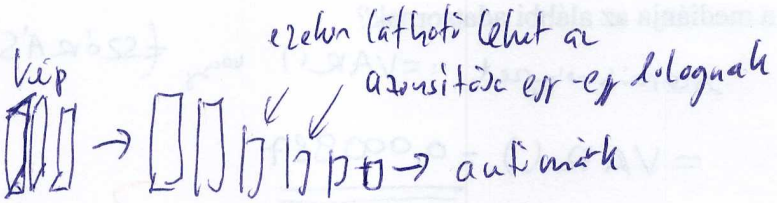
kép mérete: $W \times H \times 3$ színcsatorna (piros, zöld, kék)

rengeteg súly → rengeteg súly



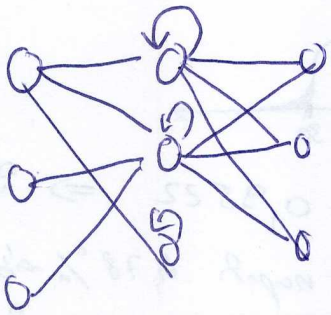
+pooling



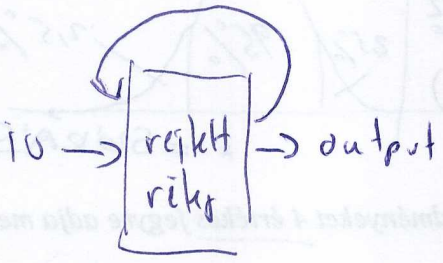


További hálók, pl.:

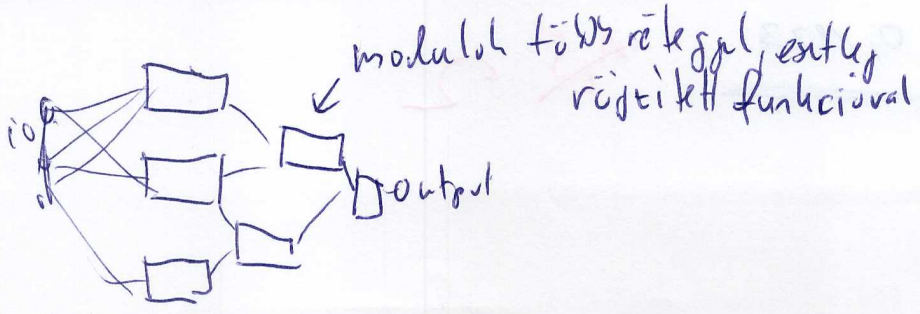
rekurrens hálók



long-short term memory



moduláris rendszerek



Modellezés Gauss eloszlásokkal:

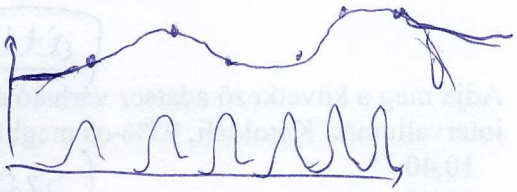
Korábban k-környezeti módszer → spam, nemspam

távolság figyelembevételével:

$$\hat{y} = \sum_{i=1}^{N_{tr}} \frac{K_n(X - X_i)\, y_i}{k_n(X - X_i)}$$

↑ sűrűségfv. pl Gauss

Igaz kivetített ábra:



problémák többdimenzióban:

különböző σ-kell (Mah.transz.!)

egyenértékűségek

1 ponthoz össze N_{tr}-kell (kernel trick ↓ zárt forma)

↙ előválogatás

„Gaussian Process..."

kernel mátrix ~~⊗~~ hat mátrix lin. regnél

hasonlóságokkal nem lin-ek.

$$\hat{y} = X(X^T X)^{-1} X^T y$$

↑ csak MLR-re