T061215

1a) Kriteriet är

$$J(\bar{m}_1, \bar{m}_2, \ldots, \bar{m}_K) = \sum_{k=1}^{K} \sum_{\bar{x} \in C_k} \|\bar{x} - \bar{m}_k\|^2$$

C_k : kluster k

J = summan av spridning runt respektive klustercentrum

J liten innebär att $\bar{m}_1 \ldots \bar{m}_K$ är bra representanter.

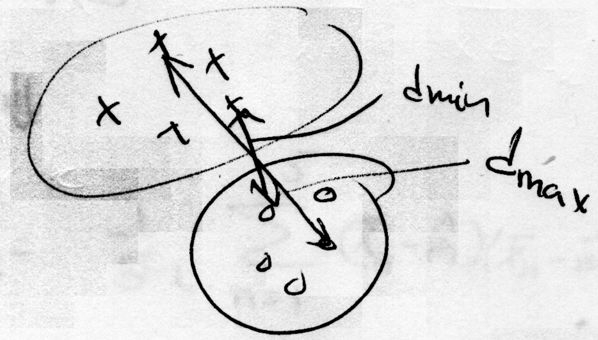
1b)

1. Beräkna distansmatris D med element D_{ij}
2. Identifiera det par (p,q) av kluster med kortast avstånd.
3. Slå ihop (p,q) till ett kluster och räkna om motsvarande avstånd till en ny distansmatris.

För att kunna räkna ut nya avstånd måste det finnas en definition för hur avstånd mellan grupper av observationer. Detta mått måste i sin tur bygga på en definition av avstånd mellan två individuella observationer

1c) Furthest neighbor = d_{max} = avståndet mellan det par av exempel som ligge längst ifrån varandra

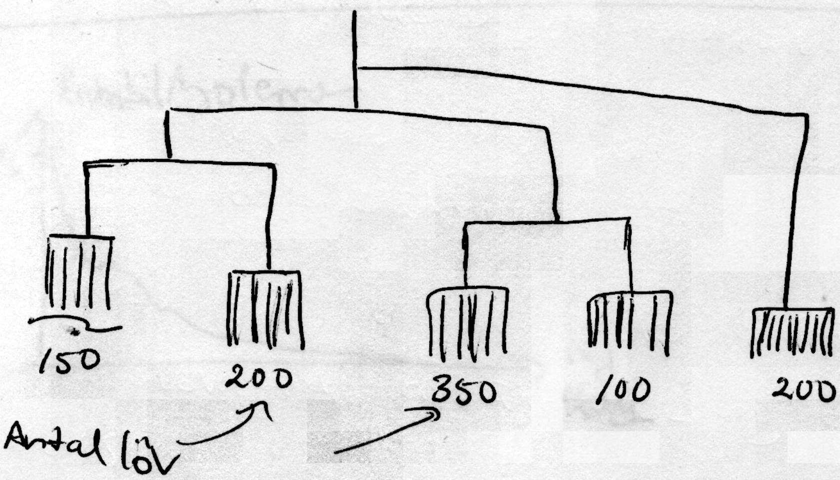
Ex.



d_{min} = Nearest neighbor = avståndet mella det par av exempel som ligge närmast varandra.

1d)

Ex.



Den hierarkiska klustringe komme föra att slå ihop alla exempel som tillhör ett klust innan nästa påbörja

1e) <u>PCA</u>: I princip har Kalle endast fem observationer $\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4, \bar{x}_5$ som motsvarar ~~...~~ var sitt kluster

$$\Downarrow$$

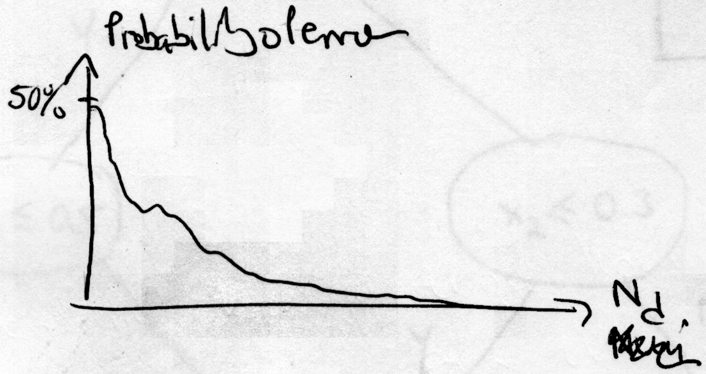
$$\underline{\underline{C}}_{xx} = \frac{1}{5-1} \sum_{n=1}^{5} (\bar{x}_n - \hat{m})(\bar{x}_n - \hat{m})$$

$$\hat{m}_x = \frac{1}{5} \sum_{n=1}^{N} \bar{x}_n$$

$$\Downarrow$$

$\underline{\underline{C}}_{xx}$ har rang 4 $\Rightarrow$ 6 egenvärde är noll.

2a)

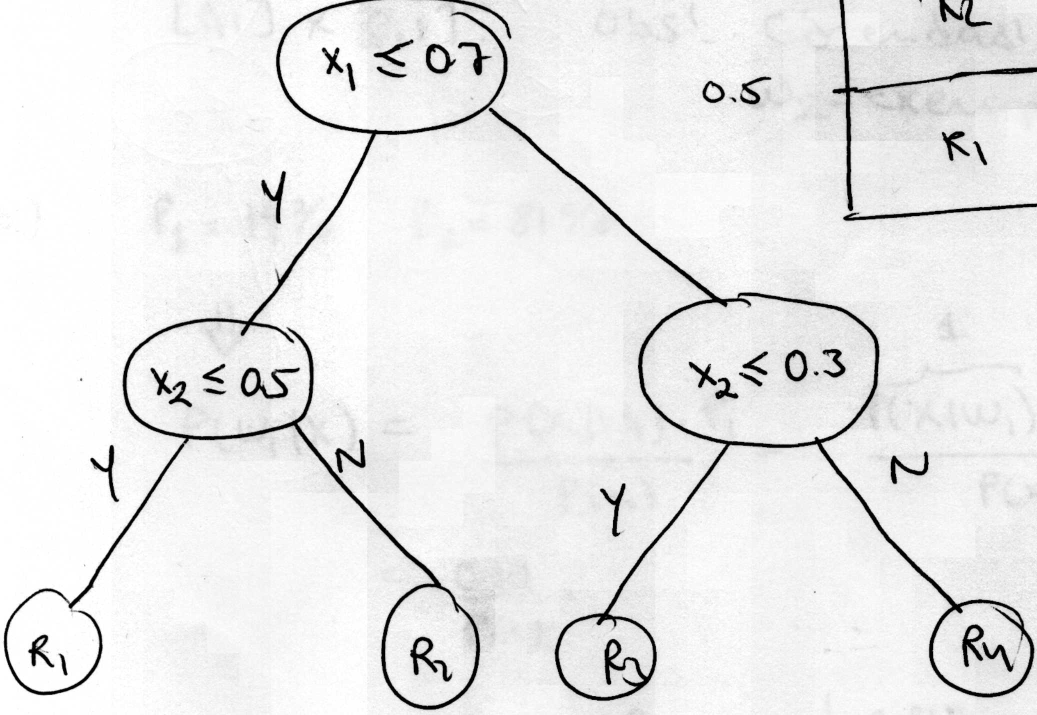


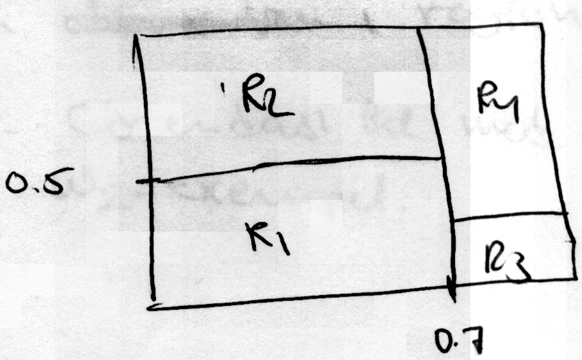
b) Linjära klassifierare överanpassar sig normalt inte lika mycket som olinjära & flexibla motsvarigheter

2c) Fisher: $\bar{y} = \underline{w}^T \bar{x}$

Linjär extraktion av särdrag $\bar{y}$ som separerar väl mellan klasserna och har relativt liten inom variation.

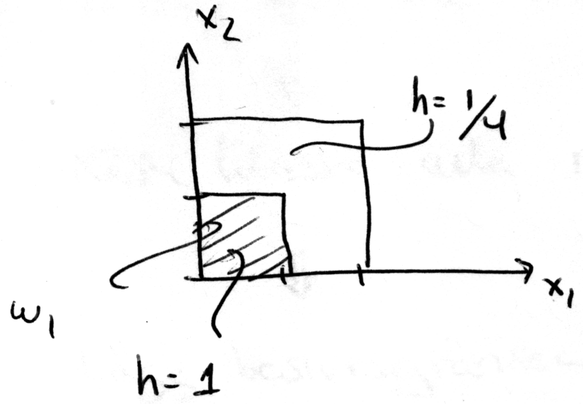
2d) (x_1, x_2)





2e) Alla beslut tas genom att studera värdet på ett särdrag i taget => Uppdelning av datarummet i rektanglar, se 2d ovan.

3a



(a) Optimalt att välja ω_1 om

$$p(\omega_1 | x) > p(\omega_2 | x)$$

$$\Downarrow$$

$$p(x | \omega_1) > p(x | \omega_2)$$

$$\Downarrow$$

Välj klass ω_1 för alla ~~observationen~~ observationer i regionen $[0,1] \times [0,1]$. Obs! Gör endast fel med ω_2-exempel.

(b) $P_1 = 19\%$ $\quad P_2 = 81\%$

$$\Downarrow$$

$$P(\omega_1 | x) = \frac{p(x | \omega_1) \cdot P_1}{p(x)} = \frac{\overbrace{p(x | \omega_1)}^{1} \cdot 0.19}{p(x)}$$

$$= \frac{0.19}{p(x)}$$

$$P(\omega_2 | x) = \frac{p(x | \omega_2) P_2}{p(x)} = \frac{\frac{1}{4} \cdot 0.81}{p(x)} = \frac{0.205}{p(x)}$$

$\Rightarrow$ Välj alltid ω_2! $\Rightarrow$ 100% fel på ω_1

(c) Extremt dyrt att klassa ω_1-exempel fel

$\Downarrow$

Måste klassa alla rätt från klass ω_1

$\Downarrow$

Lägg beslutsgränsen som i (a) så att de enda fel som ~~[illegible]~~ kommer att ske är felklassningar ~~[illegible]~~ av exempel från ω_2.

---

4 a) ~~X = B*rand(10000,~~

X = B * randn(3,10000);

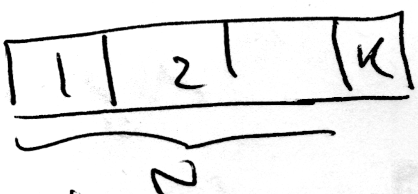
b) PCA: Hitta det underrum där varians är störst. Kolumnerna i $\underline{X}$ har ingen spridning utanför det rum som spänns av kolumnerna i $\underline{B}$.

$\Downarrow$

PCA kommer att ge ~~tre~~ egenvektorer med egenvärden som är långt ifrån noll

Dessa ~~tre~~ spänner upp samma rum som kolumnerna i $\underline{B}$.

5a)



1. Dela N exempel i K block

2. Gör design m.h.a N-1 block, testa prestanda med exemplen i det utelämnade blocket $\Rightarrow \hat{e}_k$

3. Upprepa 2 för alla K möjligheter då ett block utelämnas

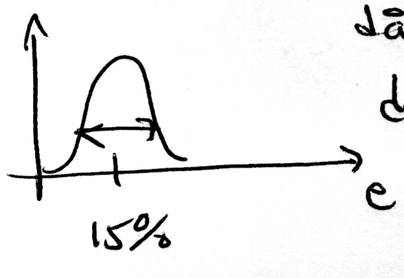
4. Beräkna estimat av prestanda genom att ta medelvärdet av de K prestanda estimaten $\hat{e}_k$

$$\hat{e}_{cv} = \frac{1}{K} \sum_{k=1}^{K} \hat{e}_k$$

5b) $\hat{e}_{cv} = 15\%$

men slutdesign $\Rightarrow$ 35% !!

Antag "prior"-fördelning $P(e|N_d)$ som är bred: Inte osannolikt att få dålig prestanda p.g.a. att designexemplen inte balanserade



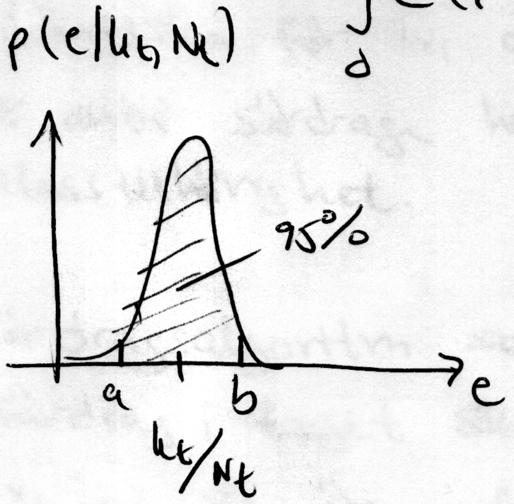
5c) Bayes sats

$$p(e|k_t,N_t) = \frac{p(k|e,N_t)\,p(e|\cancel{N_t})}{p(k_t)}$$

$$= \frac{\binom{N}{k} e^k (1-e)^{N_t-k} \cdot \overbrace{p(e)}^{=1 \text{ uniform på } [0,1]}}{p(k_t)}$$

$$\Downarrow$$

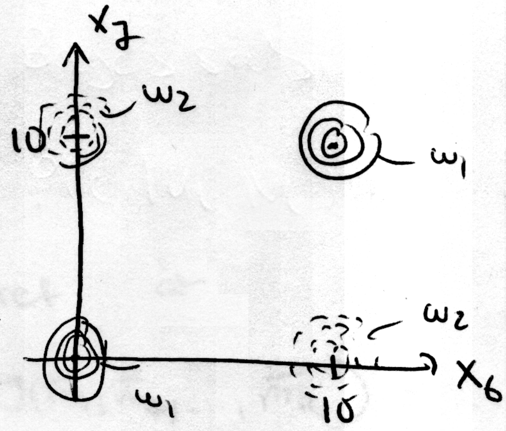
$$p(e|k_t,N_t) = \frac{e^k(1-e)^{N_t-k}}{\int_0^1 e^k(1-e)^{N_t-k}\,de}$$



Bayesianskt konfidensintervall $[a,b]$.

6

(a)



(b) Marginalfördelningarna för båda x_6 och x_7 är identiska för w_1 och w_2: Var för sig är alltså särdragen helt utan information om klasstillhörighet.

⇓

En särdragsalgoritm som endast studerar ett särdrag i taget skulle aldrig upptäcka att x_6 och x_7 är informativa tillsammans.