TO51014

1a Sv måste få tillgång till mycket stora datamängder så att hon i princip kan estimera täthetsfunktioner som beskriver data. Med få exempel blir beslutsgränserna mycket osäkra eftersom samplingen av datarummet blir mycket gles

1b Med endast 10 ~~[illegible]~~ testexempel är prestandauppskattningen mycket osäker så det är ingen signifikant skillnad att göra 2 eller 3 fel. För att minska risken för överanpassning ska alltid den enklaste modellen väljas.

1c Bra: Bättre prestanda då alla tillgängliga exempel är med för design - säkrare beslutsgräns.

Dåligt: Med så få exempel kommer prestanda på träningsdata att säga mycket lite om verklig prestanda, troligen är resultatet på träningsdata alltför optimistiskt.

1d) Korsvalidering

Ex. 5-faldig

1. | 1 | 2 | 3 | 4 | 5 |

Dela datamängde i 5 icke överlappade block.

2. Gör design m.h.a. 4 block, test med exempel i det 5:e blocket

$$\Downarrow$$

$$\hat{e}_i \quad i = 1, 2, 3, 4, 5$$

3. $$\hat{e}_{CV} = \frac{1}{5} \sum_{i=1}^{5} \hat{e}_i$$

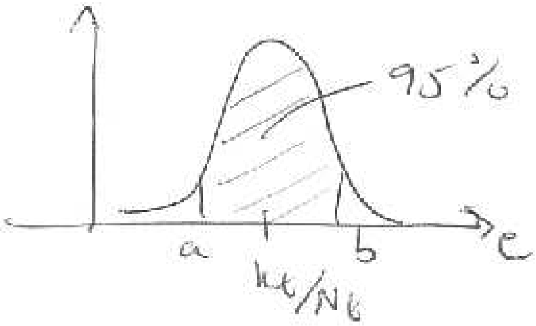
1e) $k_t = 4 \quad N_t = 20 \qquad k_t = 400 \quad N_t = 2000$

$$\hat{e} = \frac{4}{20} = \frac{400}{2000} = 20\%$$

Bayesianskt konfidensintervall

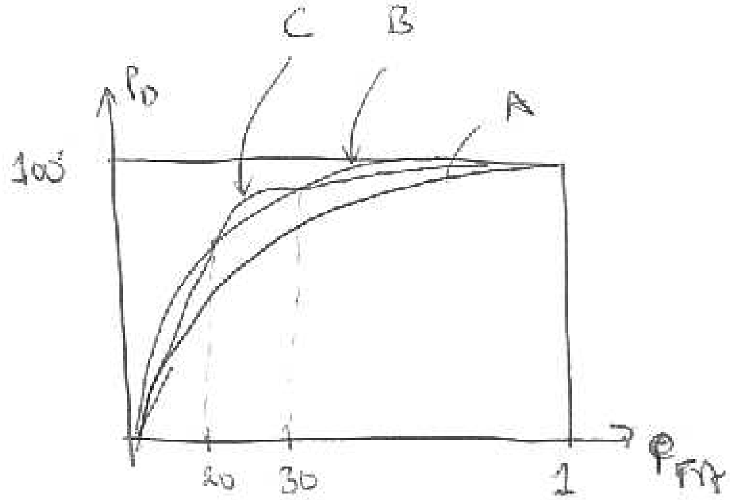
$$p(e \mid k_t, N_t) = \frac{P(k_t \mid e, N_t)\, p(e)}{P(k_t \mid N_t)} = \frac{\binom{N_t}{k_t} e^{k_t} (1-e)^{N_t - k_t}}{P(k_t \mid N_t)}$$

$$P(e \mid k_t, N_t) = \frac{e^{k_t}(1-e)^{N_t-k_t}}{\int_0^1 e^{k_t}(1-e)^{N_t-k_t}\, de}$$



$[a, b]$ = Bayesianskt konf. intervall

1f) ROC



2a SVM har större marginal eftersom perceptronalgoritmen stannar så snart alla exempel korrekt klassade

2b) Med få exempel blir det lätt överanpassning av flexibla klassificerare

2c) Fisher: Vill projicera alla $\bar{x}_n$ på en linje där de får värden

$$y_n = \hat{w}^T \bar{x}_n$$

strategin är att välja projektionsvektorn $\hat{w}$ så att

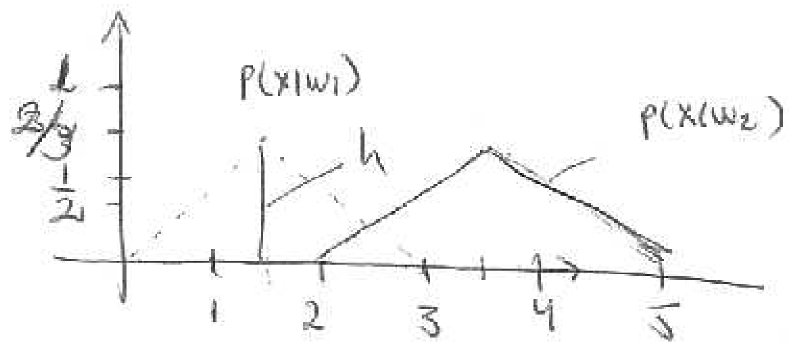
$$J(\hat{w}) = \frac{(m_1 - m_2)^2}{\sigma_1^2 + \sigma_2^2}$$

maximeras ~~minimeras~~ där $m_1 = \frac{1}{N} \sum_{y_n \in \text{klass } 1}^{N} y_n$ osv.

Med denna maximering separerar klasserna maximalt i termer av projektionsvärdena y_n

2d) Man kan tex använda ridge regression eller PLS. som ju är designade för att hantera stora korrelationer

~~2e)~~
3



$\frac{3 \cdot h}{2} = 1 \quad \boxed{h = \frac{2}{3}}$

(a) Beslutsgräns: $\{x \mid p(x|H_1) = p(x|H_2)\}$

$x = 2\frac{1}{2}$

(b) Flytta beslutsgränsen till $x=3$. Då blir inga w_1-exempel felklassade

(c) I. Måste anta/veta modell

II. Omöjligt att skatta parametrar perfekt med få data.

4a Teori för PCA $\Rightarrow \hat{x} = \bar{m} + \sum_{i=1}^{N} t_i \cdot \bar{e}_i$

ger minimal fel i minsta kvadratmening (väntevärde)

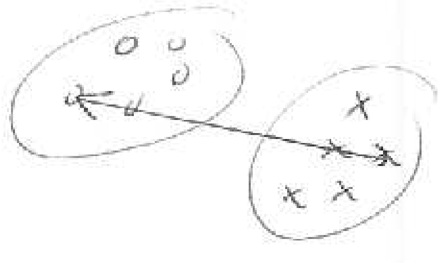
4b

$\underline{C}$: 10x10

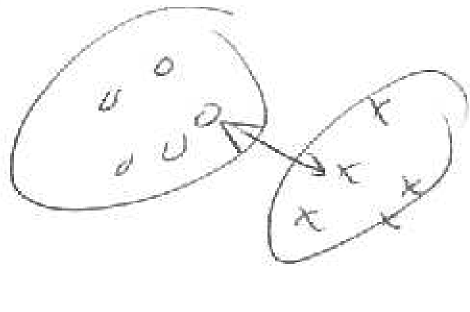
$\bar{x} \in \mathbb{R}^{10\times1}$ $\Rightarrow$ Alla exempel ligger i ett underrum av dimension 2.

$\Rightarrow$ Alla exempel kan enkelt visualiseras i en 2D-plot.

4c) Furthest neighbour complete linkage



Nearest neighbour single linkage



4 d) 1. Klustra exemplen

2. För alla exempel i ett kluster, utför PCA.

om det räcker med t.ex. 3 principalkomponenter för alle kluster kan alla exempel bestämmas med ett heltal och tre koordinater.