# Løsningsforslag eksamen 2012H

## Oppgave 1

Se lærebok og forelesningsnotater.

## Oppgave 2

Univariat toklasseproblem med tetthets-funksjonene:

$$p(x|\omega_i) = \frac{1}{\sqrt{2\pi}\,\sigma_i} e^{-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2} \quad \text{der} \begin{cases} \mu_1 = 0 & \sigma_1 = 1 \\ \mu_2 = 2 & \sigma_2 = 1 \end{cases}$$

a) Mulige diskriminantfunksjoner:

$$g_i(x) = \ln P(\omega_i|x) = \ln p(x|\omega_i) + \ln P(\omega_i) - \cancel{\ln p(x)}$$

Som gir:

$$g_i(x) = -\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2 - \frac{1}{2}\ln 2\pi - \ln\sigma_i + \ln P(\omega_i)$$

Med like à priori sannsynligheter:

$$\left.\begin{aligned} g_1(x) &= -\frac{x^2}{2} \\ g_2(x) &= -\frac{(x-2)^2}{2} \end{aligned}\right\} \Rightarrow g(x) = -\frac{x^2}{2} + \frac{(x-2)^2}{2} = -2x + 2$$

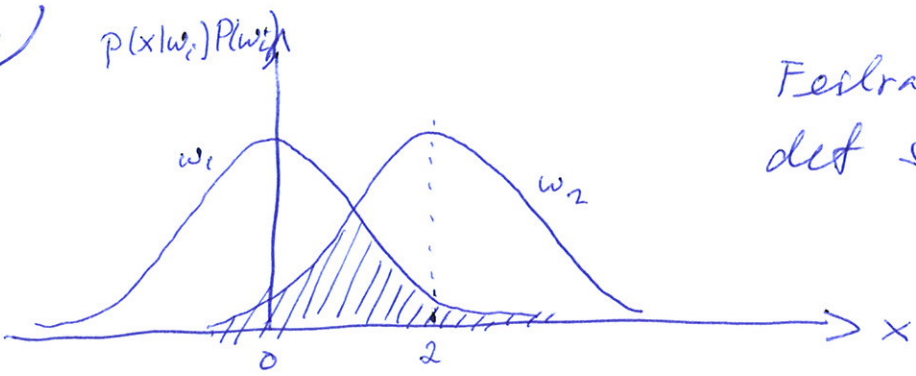
Siden kun fortegnet har betydning for valg av klasse, kan vi dele på 2 her. Toklasse diskriminantfunksjonen blir da:

$$g(x) = -x + 1$$

slik at man skal velge $\omega_1$ når $g(x) > 0$ og $\omega_2$ ellers. Beslutningsregelen kan da skrives som:

$$\boxed{\text{Velg } \omega_1 \text{ dersom } x < 1,\ \omega_2 \text{ ellers.}}$$

b)




Feilraten er gitt ved det skraverte arealet,

c) Her er $P(\omega_1) \neq P(\omega_2)$ slik at siste leddet i diskriminantfunksjonene fra pkt. a må beholdes:

$$g_i(x) = -\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2 - \frac{1}{2}\ln 2\pi - \ln\sigma_i + \ln P(\omega_i)\ , \quad i=1,2$$

Innsetting for $\mu_i$, $\sigma_i$ og $P(\omega_i)$ gir:

$$g_1(x) = -\frac{x^2}{2} + \ln\frac{2}{3} \quad \text{og} \quad g_2(x) = -\frac{(x-2)^2}{2} + \ln\frac{1}{3}$$

slik at to-klasse diskriminantfunksjonen blir:

$$g(x) = g_1(x) - g_2(x) = \underline{\underline{-2x + 2 + \ln 2}}$$

Desisjonsgrensen $x_0$ er gitt ved $g(x) = 0$ som gir:

$$x_0 = \frac{\ln 2 + 2}{2} \approx \underline{\underline{1,35}}$$

Desisjonsregelen blir da:

$$\underline{\underline{\text{Velg } \omega_1 \text{ hvis } x < x_0\ ,\ \omega_2 \text{ ellers}}}$$

d) Her er kostnadene $\lambda(\alpha_1|\omega_2) = 0{,}8$ og $\lambda(\alpha_2|\omega_1) = 0{,}2$ mens $\lambda(\alpha_1|\omega_1) = \lambda(\alpha_2|\omega_2) = 0$ der $\alpha_i$ er handlingen som består i å velge klasse $\omega_i$, der $i = 1,2$.
Betinget risk for hver av handlingene blir da:

$$R(\alpha_1|x) = \lambda(\alpha_1|\omega_1)P(\omega_1|x) + \lambda(\alpha_1|\omega_2)P(\omega_2|x)$$
$$= 0{,}0 \cdot \frac{p(x|\omega_1)P(\omega_1)}{p(x)} + 0{,}8\,\frac{p(x|\omega_2)P(\omega_2)}{p(x)}$$
$$= \underline{8p(x|\omega_2)/30 \cdot p(x)}$$

og

$$R(\alpha_2|x) = \lambda(\alpha_2|w_1)P(w_1|x) + \lambda(\alpha_2|w_2)P(w_2|x)$$

$$= 0{,}2\,\frac{p(x|w_1)P(w_1)}{p(x)} + 0 \cdot \frac{p(x|w_2)P(w_2)}{p(x)}$$

$$= 0{,}2\,\frac{p(x|w_1)\frac{2}{3}}{p(x)} = \underline{4p(x|w_1)/30p(x)}$$

Desisjonsregelen blir da:

Velg $w_1$ dersom $R(\alpha_1|x) < R(\alpha_2|x)$ og $w_2$ ellers

$\Downarrow$

— " — $\dfrac{8p(x|w_2)}{30 \cdot p(x)} < \dfrac{4p(x|w_1)}{30 \cdot p(x)}$ — " —

$\Downarrow$

— " — $2p(x|w_2) < p(x|w_1)$ — " —

$\Downarrow$

— " — $\dfrac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-2)^2} < \dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ — " —

$\Downarrow$

— " — $-\dfrac{(x-2)^2}{2} + \ln 2 < -\dfrac{x^2}{2}$ — " —

$\Downarrow$

— " — $-(x^2-4x+4) + 2\ln 2 < -x^2$ — " —

$\Downarrow$

— " — $4x - 4 + 2\ln 2 < 0$ — " —

$\Downarrow$

<u>Velg $w_1$ dersom $x < x_0$ og $w_2$ ellers</u>

der desisjonsgrensen $x_0$ er:

$$\underline{\underline{x_0 = \frac{2 - \ln 2}{2} \approx 0{,}65}}$$

# Oppgave 3

a) $g_i(\bar{x}) = \bar{w}_i^t \bar{x} + w_{i0}$, $i = 1, \dots, c$

$\bar{w}_i$ = vektvektor for klasse $\omega_i$

$w_{i0}$ = skalarvekt — " — $\omega_i$

Beslutningsregel:

Velg $\omega_m$ dersom $g_m(\bar{x}) = \max_{j} \{g_j(\bar{x})\}$ $j = 1, \dots, c$

Man velger klassen med størst verdi på diskriminant-funksjonen.

b) For to klasser:

$$g(\bar{x}) = g_1(\bar{x}) - g_2(\bar{x}) = \bar{w}^t \bar{x} + w_0$$

der $\bar{w} = \bar{w}_1 - \bar{w}_2$ og $w_0 = w_{10} - w_{20}$.

Her blir beslutningsregelen:

$$\boxed{\text{Velg } \omega_1 \text{ dersom } g(\bar{x}) > 0, \ \omega_2 \text{ ellers}}$$

c) Utvidede vektorer:

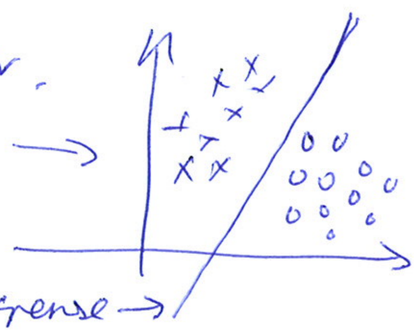
$$\bar{y} = \begin{bmatrix} 1 \\ \bar{x} \end{bmatrix} \text{ (egenskapsvektor) og } \bar{a} = \begin{bmatrix} w_0 \\ \bar{w} \end{bmatrix} \text{ (vektvektor)}$$

Kan da skrive:

$$\underline{g(\bar{x}) = \bar{a}^t \bar{y}}$$

d) Se lærebok og forelesningsnotater.

Lineært separabelt sett med $d=2$. →



e) Et eksempel er Perceptron-algoritmen:

$$\bar{a}_1 = \text{vilkårlig startvektor}$$

$$\bar{a}_{k+1} = \bar{a}_k + \rho_k \sum_{\bar{y} \in \mathcal{Y}_k} \bar{y} \quad , k = 1, 2, \ldots$$

Se lærebok og notater for detaljer

der $\mathcal{Y}_k = \{ \bar{y} : \bar{a}_k^t \bar{y} \leq 0 \}$

f) Fast inkrement regelen:

$$\bar{a}_1 = \text{vilkårlig startvektor}$$

$$\bar{a}_{k+1} = \bar{a}_k + \bar{y}^k \quad , k = 1, 2, \ldots$$

Syklisk gjennomgang av treningssettet.

Her er $\bar{y}^k$ et sample som er feilklassifisert av $\bar{a}_k$.

## Oppgave 4

a) Se lærebok og notater.

b) Blandingstetthet for $c = 2$:

$$p(\bar{x} | \bar{\theta}) = P(\omega_1) p(\bar{x} | \omega_1, \bar{\theta}_1) + P(\omega_2) p(\bar{x} | \omega_2, \bar{\theta}_2)$$

c) Likelihoodfunksjonen kan som vanlig skrives som:

$$p(\bar{x} | \bar{\theta}) = \prod_{k=1}^{n} p(\bar{x}_k | \bar{\theta})$$

der $\bar{x}_k$, $k = 1, \ldots, n$ er treningssamplene.

Log-likelihoodfunksjonen blir da:

$$\mathcal{L}(\bar{\theta}) = \sum_{k=1}^{n} \ln p(\bar{x}_k|\bar{\theta})$$

og gradienten mht. $\bar{\theta}_i$ blir:

$$\nabla_{\bar{\theta}_i} \mathcal{L}(\bar{\theta}) = \sum_{k=1}^{n} \nabla_{\bar{\theta}_i} \ln p(\bar{x}_k|\bar{\theta})$$

$$= \sum_{k=1}^{n} \frac{1}{p(\bar{x}_k|\bar{\theta})} \nabla_{\bar{\theta}_i} p(\bar{x}_k|\bar{\theta})$$

$$= \sum_{k=1}^{n} \frac{1}{p(\bar{x}_k|\bar{\theta})} \nabla_{\bar{\theta}_i} \left[ \sum_{j=1}^{2} P(\omega_j) p(\bar{x}|\omega_j, \bar{\theta}_j) \right]$$

$$= \sum_{k=1}^{n} \frac{P(\omega_i)}{p(\bar{x}_k|\bar{\theta})} \nabla_{\bar{\theta}_i} p(\bar{x}|\omega_i, \bar{\theta}_i)$$

Bruker Bayes regel:

$$P(\omega_i|\bar{x}_k, \bar{\theta}) = \frac{p(\bar{x}_k|\omega_i, \bar{\theta}_i) P(\omega_i)}{p(\bar{x}_k|\bar{\theta})}, \quad i=1,2$$

og kan da skrive:

$$\nabla_{\bar{\theta}_i} \mathcal{L}(\bar{\theta}) = \sum_{k=1}^{n} \frac{P(\omega_i|\bar{x}_k, \bar{\theta})}{p(\bar{x}_k|\omega_i, \bar{\theta}_i)} \nabla_{\bar{\theta}_i} p(\bar{x}_k|\omega_i, \bar{\theta}_i)$$

$$= \sum_{k=1}^{n} P(\omega_i|\bar{x}_k, \bar{\theta}) \nabla_{\bar{\theta}_i} \ln p(\bar{x}_k|\omega_i, \bar{\theta}_i)$$

Ved å sette gradienten lik null får man da:

$$\sum_{k=1}^{n} P(\omega_i|\bar{x}_k, \bar{\theta}) \nabla_{\bar{\theta}_i} \ln p(\bar{x}_k|\omega_i, \bar{\theta}_i) = 0, \quad i=1,2$$

d)

Her er $\bar{\theta} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ der $\mu_1$ og $\mu_2$ er ukjente.
Videre er:

$$p(x_k | w_i, \mu_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_k - \mu_i)^2} \quad \text{for } i = 1,2$$

og derav:

$$\ln p(x_k | w_i, \mu_i) = -\frac{1}{2}(x_k - \mu_0)^2 - \frac{1}{2}\ln(2\pi)$$

slik at:

$$\nabla_{\mu_i} \ln p(x_k | w_i, \mu_i) = x_k - \mu_i \quad , \quad i = 1,2$$

Innsetting i likningssystemet gir da:

$$\sum_{k=1}^{n} P(w_i | x_k, \bar{\theta})(x_k - \mu_i) = 0 \quad , \quad i = 1,2$$

$$\Downarrow$$

$$\sum_{k=1}^{n} P(w_i | x_k, \bar{\theta}) x_k = \sum_{k=1}^{n} P(w_i | \bar{x}_k, \hat{\theta}) \mu_i \quad , \quad i = 1,2$$

$$\Downarrow$$

$$\hat{\mu}_i = \frac{\sum_{k=1}^{n} P(w_i | x_k, \bar{\theta}) x_k}{\sum_{k=1}^{n} P(w_i | x_k, \bar{\theta})} \quad (\text{vektet sum av } x_k)$$

der

$$P(w_i | x_k, \bar{\theta}) = \frac{p(x_k | w_i, \mu_i) P(w_i)}{\sum_{j=1}^{2} p(x_k | w_j, \mu_j) P(w_j)} \quad \text{der } P(w_1) = \frac{1}{2} \text{ og } P(w_2) = \frac{1}{2}$$

$$= \frac{e^{-\frac{1}{2}(x_k - \mu_i)^2}}{e^{-\frac{1}{2}(x_k - \mu_1)^2} + e^{-\frac{1}{2}(x_k - \mu_2)^2}} \quad , \quad i = 1,2$$

Dette er et implisitt likningssystem for $\mu_1$ og $\mu_2$, som kan løses ved iterasjon!

e)

Her vil $P(\omega_i | x_k, \bar{\theta}) \approx \begin{cases} 1 & \text{når } x_k \in \omega_i \\ 0 & \text{når } x_k \notin \omega_i \end{cases}$

Estimatene for forventningene blir da:

$$\hat{\mu}_i \approx \frac{\sum_{x_k \in \omega_i} x_k}{n} = \underline{\underline{m_i}}$$ dvs. samplemiddelet for hver av klassene.

## Oppgave 5

a) Se lærebok og notater.

b) Den nærmeste naboen er samplet $\begin{bmatrix} 1.8 \\ 1.0 \end{bmatrix}$ fra $\omega_1$

Nærmeste-nabo regelen velger da $\underline{\underline{\omega_1}}$

De tre nærmeste naboene er

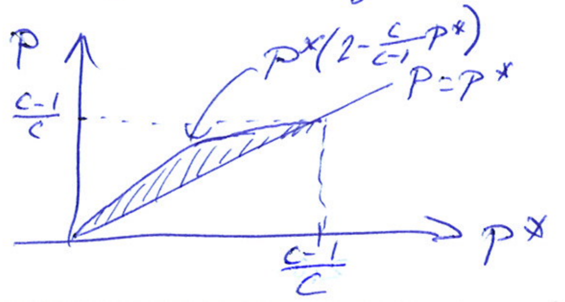
$\begin{bmatrix} 1.8 \\ 1.0 \end{bmatrix}$ fra $\omega_1$ (selvsagt) og $\begin{bmatrix} 2.1 \\ 1.5 \end{bmatrix}$ og $\begin{bmatrix} 2.5 \\ 0.8 \end{bmatrix}$ fra $\omega_3$

klasse $\omega_3$ er da i flertall slik at k-NNR med $k = 3$ velger klasse $\underline{\underline{\omega_3}}$.

c)

$$P^* \leq P \leq P^* \left( 2 - \frac{c}{c-1} P^* \right)$$

Her er $P$ = asymptotisk feilrate for NNR og $P^*$ = optimal feilrate.




Den asymptotiske feilraten vil ligge i det skraverte arealet, mellom øvre og nedre grense.