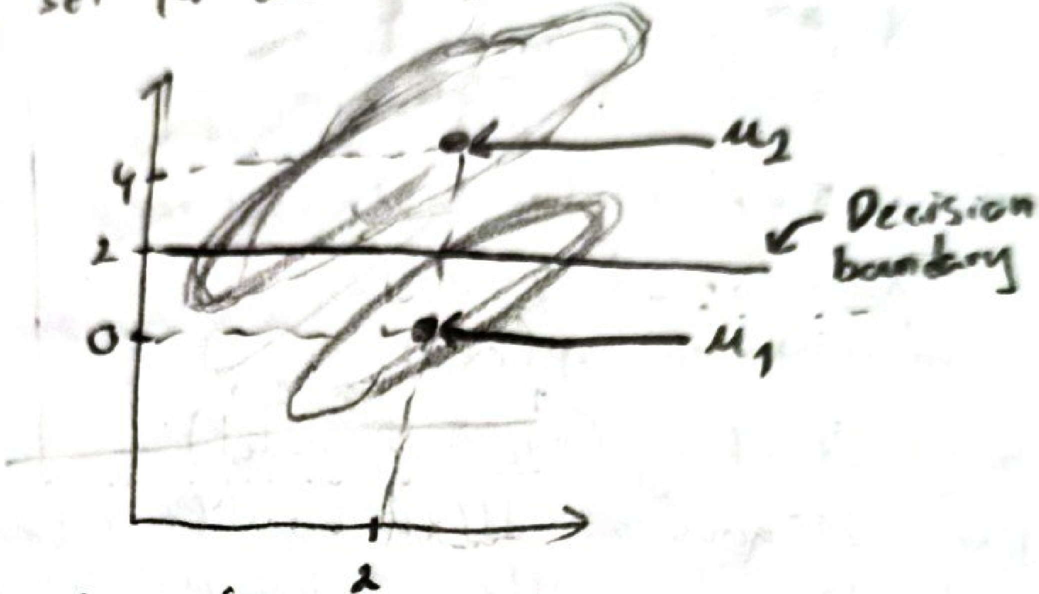
EN 2018

1

(a) Works by classifying a sample as the class of the nearest prototype.

Training: Compute prototypes (often mean vector of the training set for each class

(b) (i).



The boundary of Euclidean MDC is perpendicular to the segment connecting the means. (in the middle)

Not every sample is correctly classified by this

(ii).

$$g_k(x) = \mu_k^T C^{-1} x - 0.5 \mu_k^T C^{-1} \mu_k$$

$$g_1(x) = (2\ 0)\begin{pmatrix} 1 & -2 \\ -2 & 5 \end{pmatrix} x - 0,5 (2\ \ 0) \begin{pmatrix} 1 & -2 \\ -2 & 5 \end{pmatrix} \begin{pmatrix} 2 \\ 0 \end{pmatrix} =$$

$$= (2\ \ -4)x - 2 \overset{/2}{\Rightarrow} (1\ \ -2)x - 1$$

$$g_2(x) = (-6\ \ 16)x - 52 \overset{/2}{\Rightarrow} (-3\ \ 8)x - 26$$

$$\left.\begin{aligned} g_1\left(\begin{pmatrix} -2 \\ 0 \end{pmatrix}\right) &= -2 - 1 = -6 \\ g_2\left(\begin{pmatrix} -2 \\ 0 \end{pmatrix}\right) &= 6 - 26 = -20 \end{aligned}\right\} \text{Pertence a } \omega_1$$

(iii) $d_{ij}(x) = (\mu_i - \mu_j)^T C^{-1} (x - 0.5(\mu_i - \mu_j))$

$\mu_1 - \mu_2 = \begin{pmatrix} 0 \\ -4 \end{pmatrix}$

$b = 0.5(\mu_1 + \mu_2) = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$

$w = (\mu_1 - \mu_2)^T C^{-1} = (0 \quad -4) \begin{pmatrix} 1 & -2 \\ -2 & 5 \end{pmatrix} = (8 \quad -20)$

$d_{12} = w(x - b) = (8 \quad -20) \begin{pmatrix} x_1 - 2 \\ x_2 - 2 \end{pmatrix} =$

$= 8(x_1 - 2) - 20(x_2 - 2) =$

$= 8x_1 - 16 - 20x_2 + 40 =$ $\underbrace{}_{24}$

$= \underbrace{8x_1 - 20x_2 + 24 \Rightarrow 2x_1 - 5x_2 + 6}_{\text{Decision boundary}}$

The decision hyperplane is at $d_{12}(x) = 0 \Leftrightarrow$

$\Leftrightarrow 2x_1 - 5x_2 + 6 = 0 \Leftrightarrow x_2 = \dfrac{-2x_1 - 6}{-5} \Leftrightarrow$

$\Leftrightarrow x_2 = \dfrac{2}{5}x_1 + \dfrac{6}{5}$

This is a linear hyperplane, which makes sense since $C_1 = C_2$

② ⓐ A dendrogram is a tree diagram that records the sequence of merges (or splits) in a hierarchical clustering procedure.

ⓑ

Lance-Williams: (used to update the distance matrix)

$$d_{(i \cup j),k} = \alpha_i d_{ik} + \alpha_j d_{jk} + \beta d_{ij} + \gamma \lvert d_{ik} - d_{jk} \rvert$$

for Ward's method:

$$\alpha_i = \frac{n_i + n_k}{n_i + n_j + n_k} \; ; \; \alpha_j = \frac{n_j + n_k}{n_i + n_j + n_k} \; ; \; \beta = -\frac{n_k}{n_i + n_j + n_k} \; ; \; c = 0$$

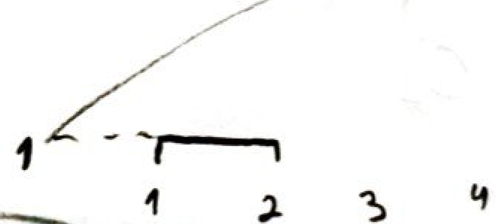
n_i, n_j, n_k are the num of elements in clusters $i, j,$ and k

1 - Initial matrix

$$\begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{array} \begin{array}{c} \begin{array}{cccc} P_1 & P_2 & P_3 & P_4 \end{array} \\ \begin{pmatrix} 0 & 1 & 4 & 5 \\ & 0 & 2 & 6 \\ & & 0 & 3 \\ & & & 0 \end{pmatrix} \end{array}$$

Smallest distance is $d_{12} = 1$ merge them

$C_1 = P_1 \cup P_2$



Size of clusters: $n_{C_1} = 2, \; n_{P_3} = 1; \; n_{P_4} = 1$

Compute new distances

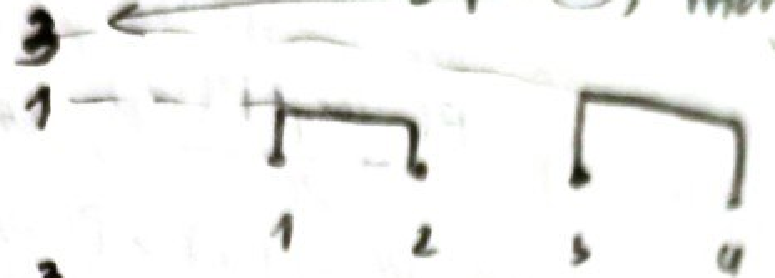
$$d_{C_1P_3} = \frac{1+1}{1+1+1} \times 4 + \frac{1+1}{3} \times 2 - \frac{1}{3} \times 1 = \frac{8}{3} + \frac{4}{3} - \frac{1}{3} \approx 3{,}67$$

$$d_{C_1P_4} = \frac{2}{3} \times 5 + \frac{2}{3} \times 6 - \frac{1}{3} \times 1 = \frac{10 + 12 - 1}{3} = 7$$

2 - Second matrix

$$\begin{array}{c|ccc} & C_1 & P_3 & P_4 \\ \hline C_1 & \multicolumn{3}{c}{\begin{pmatrix} 0 & 3{,}67 & 7 \\ & 0 & 3 \\ & & 0 \end{pmatrix}} \\ P_3 & & & \\ P_4 & & & \end{array}$$

smallest dist is $d_{34} = 3$, merge them



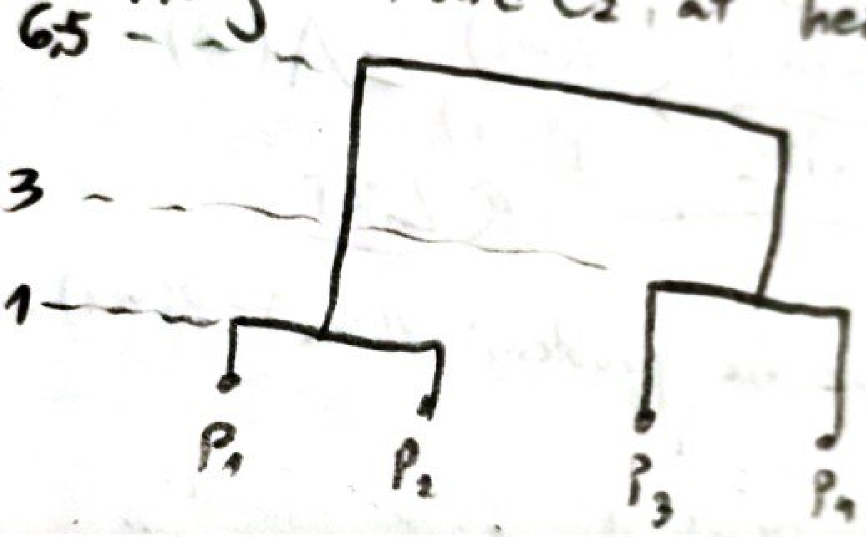
$C_2 = P_3 \cup P_4$

$d_{C_2C_1} = \frac{3}{4} 3{,}67 + \frac{3}{4} 7 - \frac{2}{4} 3 \approx 6{,}5$

$n_{C_1} = 2$
$n_{P_3} = 1$
$n_{P_4} = 1$

3 -

Merge C_1 and C_2, at height 6,5



③

$C = I$

$\mu_1 = (-1 \quad -1)^T \quad \mu_2 = (1 \quad 1)^T$

Equal prevalence: $P(\omega_1) = P(\omega_2)$

ⓐ $D = 2$ because there are 2 features ?bruh?

ⓑ In Bayes decision rule, we want to pick the highest $P(\omega_i | x) = \frac{P(x|\omega_i) P(\omega_i)}{P(x)}$, to decide for ω_1, we check $P(\omega_1 | x) > P(\omega_2 | x) \Leftrightarrow$

$$\Leftrightarrow \frac{P(x|\omega_1) P(\omega_1)}{\cancel{P(x)}} > \frac{P(x|\omega_2) P(\omega_2)}{\cancel{P(x)}} \Leftrightarrow$$

we can remove $P(x)$

$$\Leftrightarrow \boxed{\frac{P(x|\omega_1)}{P(x|\omega_2)} > \frac{P(\omega_2)}{P(\omega_1)}} = \Lambda(x)$$

LRT

in this specific problem: $P(\omega_2) = P(\omega_1)$, so:

$$\Lambda(x) \doteq \frac{P(x|\omega_1)}{P(x|\omega_2)} > 1$$

ⓒ

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^T C_i^{-1}(x-\mu_i) - \frac{1}{2}\ln|C_i| + \ln P(w_i)$$

But $C_1 = I = C_2$, so $-\frac{1}{2}x^T C^{-1} x$ and $-\frac{1}{2}\ln|C|$ cancel out in:

$$d_{12}(x) = g_1(x) - g_2(x) =$$

$$= (\mu_1 - \mu_2)^T C^{-1} x - \frac{1}{2}\mu_1^T C^{-1}\mu_1 + \frac{1}{2}\mu_2^T C^{-1}\mu_2 + \ln\left(\frac{P(w_1)}{P(w_2)}\right)$$

$\ln(1) = 0$

$$= \underbrace{(\mu_1-\mu_2)^T C^{-1}}_{\underbrace{(C^{-1}(\mu_1-\mu_2))^T}_{w}} x \underbrace{- \frac{1}{2}\mu_1^T C^{-1}\mu_1 + \frac{1}{2}\mu_2^T C^{-1}\mu_2}_{b(w_0)}$$

ⓓ

$$w = I(\mu_1 - \mu_2) = \begin{pmatrix} -2 \\ -2 \end{pmatrix}$$

$b(w_0)$:

$$-\frac{1}{2}\mu_1^T I \mu_1 = (-1 \ \ -1)\begin{pmatrix} -1 \\ -1 \end{pmatrix}(-0,5) = -1$$

$$\frac{1}{2}\mu_2^T I \mu_2 = (1 \ \ 1)\begin{pmatrix} 1 \\ 1 \end{pmatrix} 0,5 = 1$$

$$w_0 = -1 + 1 = 0$$

$$d_{12}(x) = (-2 \ \ -2)\, x = -2x_1 - 2x_2$$

This is equivalent to Euclidean MDC:

1- Bayes = Mahalanobis when $P(w_i) = P(w_j)$ for all classes

2- Mahalanobis = Euclidean when $C = I$.

Here, $C_1 = C_2 = I$ and $P(w_1) = P(w_2)$

④

ⓐ $\gamma = \frac{2}{\|w\|} = \Big|_{w=[-1\ 1]^T} \frac{2}{\sqrt{2}} \simeq 1{,}414$

ⓑ OVO:

$$M = \begin{array}{c|ccc} & SVM1 & SVM2 & SVM3 \\ \hline \omega_1 & +1 & +1 & 0 \\ \omega_2 & -1 & 0 & +1 \\ \omega_3 & 0 & -1 & -1 \end{array}$$

OVA:

$$M = \begin{pmatrix} +1 & -1 & -1 \\ -1 & +1 & -1 \\ -1 & -1 & +1 \end{pmatrix}$$

ⓒ It's a sv if $|w^T x + b| = 1$ for any SVM

$g_1(x) = -6 + 1 + 1 = -4 \rightarrow -1$ is ω_2

$g_2(x) = -6 - 1 + 7 = 0 \rightarrow$ on the boundary

$g_3(x) = -1 + 2 = 1 \rightarrow$ It's a sv for SVM 3

+1 is ω_2

so x belongs to ω_2 and is SV for SVM3

codes: $\omega_1 = (+1\ \ +1\ \ 0)$
$\omega_2 = (-1\ \ 0\ \ +1)$
$\omega_3 = (0\ \ -1\ \ -1)$

(b) $x = (2\ 0)$:

$g_1(x) = -2+1 = -1 \to -1$
$g_2(x) = -2+7 = 5 \to +1$
$g_3(x) = 0+2 = 2 \to +1$

output code $(-1\ +1\ +1)$ which has Hamming distance 0 with the code for ω_2

distances to:

ω_1 (indices 1,2): $d(\overbrace{(-1\ +1)}^{out}, (1\ 1)) = 1$
ω_2 (indices 1,3): $d((-1\ +1), (-1\ +1)) = 0$
ω_3 (indices 2,3): $d((+1\ +1), (-1\ -1)) = 2$

classify as ω_2 (Missed)

$x = (3\ 0)$

$g_1(x) = -3+1 = -2 \to -1$
$g_2(x) = -3+7 = 4 \to +1$
$g_3(x) = 0+2 = 2 \to +1$

out: $(-1\ +1\ +1)$
Hd = 0 with class ω_2

classify as ω_2 (Correct)

$x = (5\ 5)$

$g_1(x) = -5+5+1 = 1 \to +1$
$g_2(x) = -5-5+7 = -3 \to -1$
$g_3(x) = -5+2 = -3 \to -1$

out: $(+1\ -1\ -1)$
Hd = 0 with class ω_3

classify as ω_3 (Correct)

66% accuracy.

(5)

```
import numpy as np

def estimate_knn_pdf(data, K):
    N = len(data)
    pdf_estimates = []           return closest K dists
    for x in data:                 ↙
        dists = get_K_closest(data, x, y)
        d_K = distances[-1] # last neighbor

        V = 2 * d_K
        if V == 0:
            prob = 0
        else
            prob = K/(N*V)
        pdf_estimates.append(prob)
    return pdf_estimates
```

$$P(w_1|x) > P(w_2|x) \iff \frac{P(x|w_1)P(w_1)}{\cancel{P(x)}} > \frac{P(x|w_2)P(w_2)}{\cancel{P(x)}}$$

$$\iff \frac{P(x|w_1)}{P(x|w_2)} > \frac{P(w_2)}{P(w_1)}$$