EE 888/EE 8209
MIDTERM Winter 2009

SOLUTIONS (PAPER A) → comments for paper B included

(Q1) (a) Feature extraction is the process of transforming the raw input data into some reduced representation. The representation thus reflects some measurable property or subset of the raw data – preferably chosen so as to facilitate discrimination between known classes.

– desireable properties
- Invariance (rotation, scaling, etc).
- Good discrimination/sep. between classes
- Compact/low variance within class

(b) Generalization: ability for classifier to perform well on unseen/novel patterns

TRADEOFF ↔

Overfitting. ability for classifier to perform well on training data – if highly fitted, may do well in classifying training samples at the expense of unseen novel samples.

(c) BDT vs LDA.
In BDT we assume the FORM of the distribution probability of each class, through which we attain an implicit boundary between classes. In LDA we do not assume anything about the form of the distribution of each class, but rather we assume the explicit form of the boundary itself. In LDF for instance, the form used to model the boundary is linear $W^T x + w_0 = g(x)$.

(d) $\mu_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ $\mu_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ $\Sigma_1 = \begin{pmatrix} 2 & 3 \\ 3 & 2 \end{pmatrix}$ $x_1 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$

* decision rule : choose ω_1 if $g_1(x) > g_2(x)$
otherwise choose ω_2.

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^T \Sigma_1^{-1} (x-\mu_i)$$

$$\underbrace{-\frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_1| + \ln P(\omega_i)}_{\text{constant}}$$

since $\Sigma_i = \Sigma_j = \Sigma_1$ and $P(\omega_1) = P(\omega_2)$

$$\therefore g_i(x) = -\frac{1}{2}(x-\mu_i)^T \Sigma_1^{-1} (x-\mu_i)$$

$$|\Sigma_1| = 2(2) - 3(3) = -5$$

$$\Sigma_1^{-1} = -\frac{1}{5}\begin{pmatrix} 2 & -3 \\ -3 & 2 \end{pmatrix} = \begin{pmatrix} -2/5 & 3/5 \\ \frac{3}{5} & -2/5 \end{pmatrix}.$$

$$(x_1 - \mu_1) = \begin{pmatrix} 0 \\ -1 \end{pmatrix} - \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$$

$$(x_1 - \mu_2) = \begin{pmatrix} 0 \\ -1 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ -2 \end{pmatrix}.$$

$$\therefore g_1(x_1) = \frac{-1}{2}(1 \;\; -2)\begin{pmatrix} -2/5 & 3/5 \\ 3/5 & -2/5 \end{pmatrix}\begin{pmatrix} 1 \\ -2 \end{pmatrix}$$

$$= \frac{-1}{2}\left(\left(-2/5 - 6/5\right) \left(3/5 + 4/5\right)\right)\begin{pmatrix} 1 \\ -2 \end{pmatrix} = \frac{-1}{2}\left(-8/5 \;\; 7/5\right)\begin{pmatrix} 1 \\ -2 \end{pmatrix}$$

$$= -\frac{1}{2}\left(\frac{-22}{5}\right) = \frac{11}{5}$$

$$g_2(x_1) = -\frac{1}{2}\begin{pmatrix}-1 & -2\end{pmatrix}\begin{pmatrix}-2/5 & 3/5 \\ 3/5 & -2/5\end{pmatrix}\begin{pmatrix}-1 \\ -2\end{pmatrix}$$

$$= -\frac{1}{2}\begin{pmatrix}\left(2/5 - 6/5\right) & \left(-3/5 + 4/5\right)\end{pmatrix}\begin{pmatrix}-1 \\ -2\end{pmatrix}$$

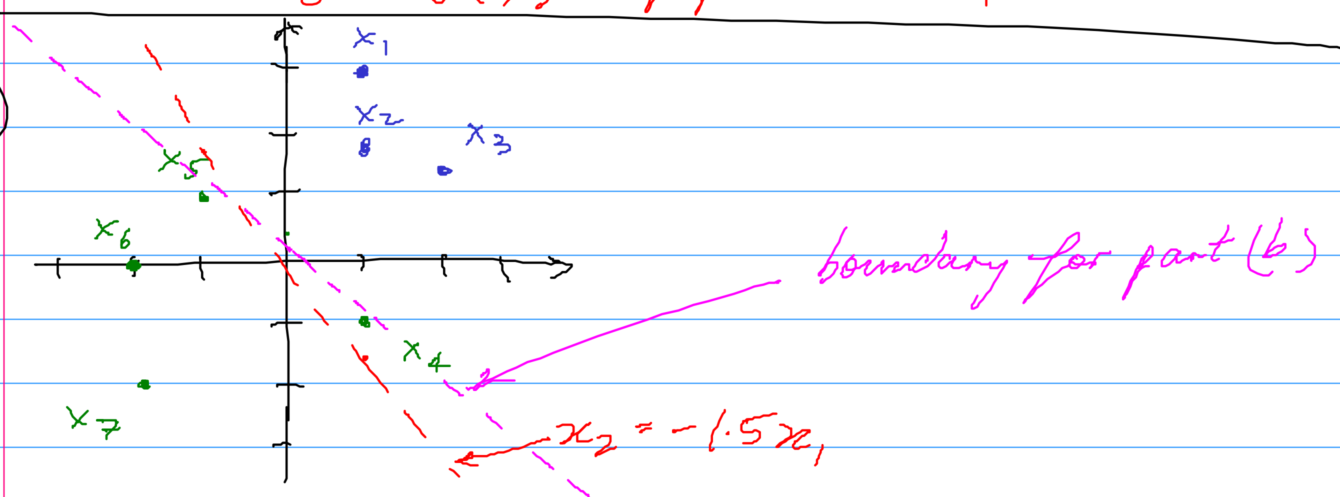
$$= -\frac{1}{2}\begin{pmatrix}-4/5 & 1/5\end{pmatrix}\begin{pmatrix}-1 \\ -2\end{pmatrix} = -\frac{1}{2}\left(2/5\right) = -1/5$$

since $g_1(x_1) > g_2(x_1)$

$$\Rightarrow \boxed{x_1 \in \omega_1}$$

* $g_2(x_1) > g_1(x_1)$ for paper B $\rightarrow x_1 \in \omega_2$

2



(a) $x_1 = (1 \;\; 3) \rightarrow \omega_1$ $\qquad g(x_1) = 4.5$

$x_7 = (-2 \;\; -2) \rightarrow \omega_2$ $\qquad g(x_7) = -5$

$g(0) = 0$.

$$g(x) = w_0 + w_1 x_1 + w_2 x_2$$

$$g(0) = w_0 = 0$$

$$\therefore \quad g(x_1) = w_1(1) + 3w_2 = 4.5 \rightarrow w_1 = -3w_2 + 4.5$$

$$g(x_7) = -2w_1 - 2w_2 = -5$$

$$-2(-3w_2 + 4.5) - 2w_2 = -5$$

$$4w_2 = 4 \qquad \Rightarrow w_1 = -3 + 4.5$$

$$w_2 = 1 \qquad\qquad = 1.5.$$

$$\therefore \quad a = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1.5 \\ 1 \end{bmatrix}$$

$$\therefore \quad g(x) = a^T y = [0 \;\; 1.5 \;\; 1] \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix}$$

$$= 1.5x_1 + x_2$$

$$1.5x_1 + x_2 = 0$$

$$x_2 = -1.5x_1$$

$\Rightarrow$ boundary classifies x_4 incorrectly $\quad x_4 = (1 \;\; -1)$

$1.5(+1) + -1 = 0.5$ should be negative!

(b). $x_1 \rightarrow 1 \;\; 3 \quad w_1$

$x_2 \rightarrow 1 \;\; 2 \quad w_1$

$x_5 \rightarrow -1 \;\; 1 \quad w_2$

$x_4 \rightarrow 1 \;\; -1 \quad w_2$

$\Rightarrow$ augment & normalize

$$\hat{Y} = \begin{array}{cccc} x_1 & x_2 & x_5 & x_4 \\ \hat{y}_1 & \hat{y}_2 & \hat{y}_3 & \hat{y}_4 \end{array}$$
$$\hat{Y} = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 \\ 3 & 2 & -1 & 1 \end{bmatrix}$$

single sample criteria

if $a^T \hat{y}_k < 0$

$a(k+1) \leftarrow a(k) + \eta \hat{y}_k$.

else $a(k+1) \leftarrow a(k)$

$$a(0) = \begin{pmatrix} -0.1 \\ -0.1 \\ -0.1 \end{pmatrix} \qquad \eta(\cdot) = 1$$

step 1 : $a(0)^T \tilde{y}_1$

$$= \begin{pmatrix} -0.1 & -0.1 & -0.1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}$$

$$= -0.5 < 0$$

missclassified $\therefore$ update

$$\Rightarrow \quad a(1) = \begin{pmatrix} -0.1 \\ -0.1 \\ -0.1 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 0.9 \\ 0.9 \\ 2.9 \end{pmatrix}$$

step 2

$$a(1)^T y_2 = \begin{pmatrix} 0.9 & 0.9 & 2.9 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$$

$$= 0.9 + 0.9 + 5.8$$

$$= 7.6 > 0$$

$\Rightarrow$ correctly classified $\therefore$ no change

$$\therefore \quad a(2)^T = \begin{pmatrix} 0.9 & 0.9 & 2.9 \end{pmatrix}$$

step 3 $a(2)^T y_3 = \begin{pmatrix} 0.9 & 0.9 & 2.9 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix}$

$$= -2.9 < 0$$

$\Rightarrow$ missclassified $\therefore$ update

$$\therefore \quad a(3) = \begin{pmatrix} 0.9 \\ 0.9 \\ 2.9 \end{pmatrix} + \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -0.1 \\ 1.9 \\ 1.9 \end{pmatrix}$$

step 4 $a(3)^T y_4 = \begin{pmatrix} -0.1 & 1.9 & 1.9 \end{pmatrix} \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$

$= 0.1 - 1.9 + 1.9 > 0$ $\Rightarrow$ classified correctly $\therefore$ no update.

$\therefore a(4) = \begin{pmatrix} -0.1 \\ 1.9 \\ 1.9 \end{pmatrix}$

** when you draw boundary → all samples correctly classified $\therefore$ would expect no further iterations to cause updates.

$g(x) = a^T y = \begin{pmatrix} -0.1 & 1.9 & 1.9 \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix}$

$= -0.1 + 1.9x_1 + 1.9x_2 \Rightarrow x_2 = \frac{-1.9x_1}{1.9} + \frac{0.1}{1.9}$

* same approach for paper B, just different order for presenting & testing $\hat{y}_k$'s

(c) Single sample perceptron vs LMS.

→ LMS uses every sample to update a. (updates at every step)

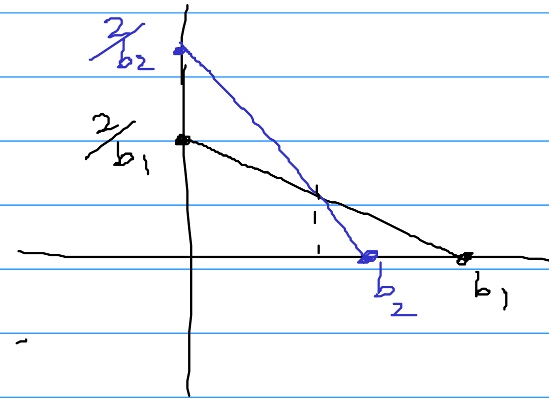
← SSP only uses sample if it is missclassified by the current a.

→ LMS imposes a margin into the update process

← SSP has no margin

(these are the main diffs)

③ $$p(x|\omega_i) = \begin{cases} 0 & x \le 0 \\ \frac{2}{b_i}\left(1 - \frac{x}{b_i}\right) & 0 \le x \le b_i \\ 0 & \text{otherwise.} \end{cases}$$



(a) $b_1 > b_2$.

$\omega_1 \rightarrow p(x|\omega_1) = \frac{2}{b_1}\left(1 - \frac{x}{b_1}\right)$

$\omega_2 \rightarrow p(x|\omega_2) = \frac{2}{b_2}\left(1 - \frac{x}{b_2}\right)$.

likelihood $\rightarrow$ $$\frac{p(x|\omega_1)}{p(x|\omega_2)} = \frac{\frac{2}{b_1}\left(1 - \frac{x}{b_1}\right)}{\frac{2}{b_2}\left(1 - \frac{x}{b_2}\right)}.$$

$$= \frac{b_2^2}{b_1^2} \frac{(b_1 - x)}{(b_2 - x)}$$

(b) min error rate classifier

$\longrightarrow$ zero-one loss $\lambda_{11} = \lambda_{22} = 0$.
$\lambda_{12} = \lambda_{21} = 1$

$\longrightarrow$ assume equal priors
$P(\omega_1) = P(\omega_2) = 0.5$

assumptions

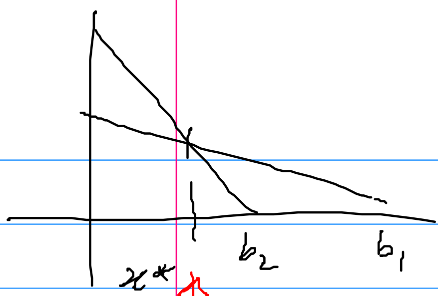
$\therefore$ likelihood fn: $$\frac{p(x|\omega_1)}{p(x|\omega_2)} > \frac{\lambda_{21} - \lambda_{11}}{\lambda_{12} - \lambda_{22}} \frac{P(\omega_1)}{P(\omega_2)}$$

decision boundary @

$$\frac{b_2^2}{b_1^2} \frac{(b_1 - x)}{(b_2 - x)} > 1$$

$$b_2^2(b_1 - x) > b_1^2(b_2 - x).$$

$$x(b_1^2 - b_2^2) > b_1 b_2(b_1 - b_2).$$

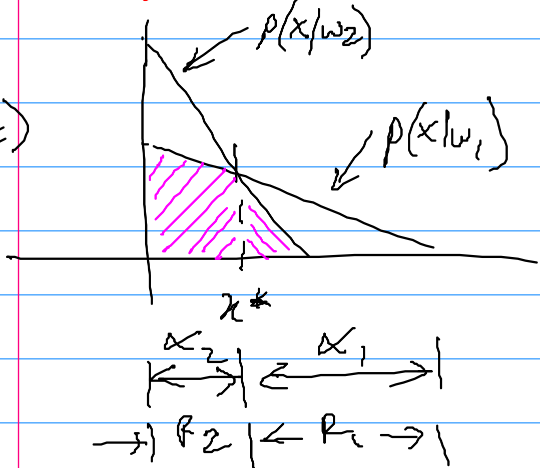


$$\frac{b_1 b_2}{b_1+b_2}$$

$$x > \frac{b_1 b_2 (b_1 - b_2)}{b_1^2 - b_2^2} \quad = (b_1+b_2)(b_1-b_2)$$

$$x > \frac{b_1 b_2}{b_1 + b_2}$$

(c)



$$P_E =$$

$$P(\text{error}) = P(\alpha_2, x \in w_1) + P(\alpha_1, x \in w_2)$$

$$= \int_{R_2} p(x|w_1) P(w_1)\,dx + \int_{R_1} p(x|w_2) P(w_2)\,dx.$$

$$P_E = \int_0^{x^*} \frac{2}{b_1}\left(\frac{b_1 - x}{b_1}\right)\cdot\frac{1}{2}\,dx + \int_{x^*}^{b_2} \frac{2}{b_2}\left(\frac{b_2 - x}{b_2}\right)\frac{1}{2}\,dx$$

$$= \frac{1}{b_1^2}\int_0^{x^*}(b_1 - x)\,dx + \frac{1}{b_2^2}\int_{x^*}^{b_2}(b_2 - x)\,dx$$

$$= \frac{1}{b_1^2}\left[b_1 x - \frac{x^2}{2}\right]_0^{x^*} + \frac{1}{b_2^2}\left[b_2 x - \frac{x^2}{2}\right]_{x^*}^{b_2}$$

$$= \frac{1}{b_1^2}\left[b_1 x^* - \frac{x^{*2}}{2}\right] + \frac{1}{b_2^2}\left[\overbrace{\left(b_2^2 - \frac{b_2^2}{2}\right)}^{=\frac{b_2^2}{2}} - \left(b_2 x^* - \frac{x^{*2}}{2}\right)\right]$$

$$= \left(\frac{1}{b_1} - \frac{1}{b_2}\right)x^* - \frac{1}{2}\left(\frac{1}{b_1^2} - \frac{1}{b_2^2}\right)x^{*2} + \frac{1}{2}.$$

$$\frac{dP_E}{dx^*} = \left(\frac{1}{b_1} - \frac{1}{b_2}\right) - \frac{2}{2}\left(\frac{1}{b_1^2} - \frac{1}{b_2^2}\right)x^* = 0.$$

$$\frac{b_2 - b_1}{b_1 b_2} - \left[\frac{b_2^2 - b_1^2}{(b_1 b_2)^2}\right] x^* = 0 .$$

$$x^* = \frac{(b_1 b_2)^2 (b_2 - b_1)}{(b_1 b_2)(b_2^2 - b_1^2)}$$

$$= \frac{b_1 b_2}{(b_2 + b_1)} \rightarrow \text{same as part (b)}$$

d) $P_E = \left(\frac{1}{b_1} - \frac{1}{b_2}\right) x^* - \frac{1}{2}\left(\frac{1}{b_1^2} - \frac{1}{b_2^2}\right) x^{*2} + \frac{1}{2}$.

$\boxed{\begin{array}{l} b_1 = 10 \\ b_2 = 3 \end{array}}$ $= \left(\frac{1}{10} - \frac{1}{3}\right)\left(\frac{30}{13}\right) - \frac{1}{2}\left(\frac{1}{100} - \frac{1}{9}\right)\left(\frac{30}{13}\right)^2 + \frac{1}{2}$

$\frac{b_1 b_2}{b_2 + b_1} = \frac{30}{13}$.

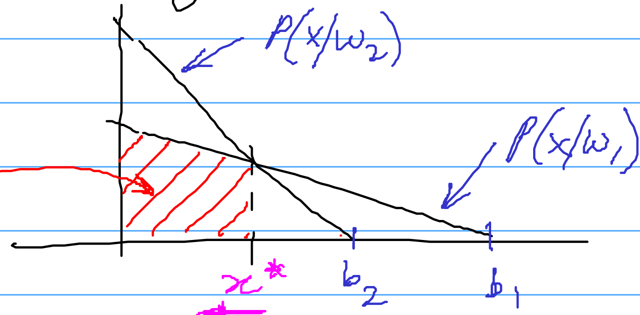
$$\boxed{P_E = 0.2308}$$

* for paper B $\rightarrow$ insert $\boxed{\begin{array}{l} b_1 = 5 \\ b_2 = 3 \end{array}}$

e) $\lambda_{21} > \lambda_{12}$

$\Rightarrow$ classifying ω_1 incorrectly more costly

(* move RIGHT for paper B.)



error in classifying x as ω_2 if $x \in \omega_1$

α_1 classify $x \in \omega_1$

to REDUCE overall error, move x^* LEFT (toward origin)