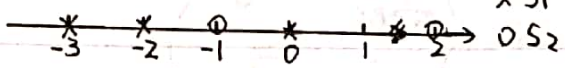
1-(a).

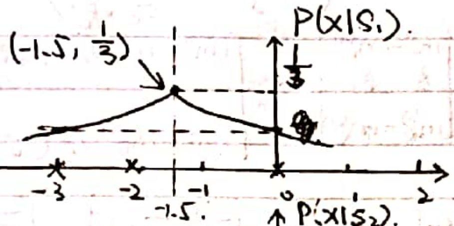


when $x \le -1.5$,

$\hat{P}(x|S_1) = \frac{k/3}{V_3} = \frac{3/3}{2|x-0|} = -\frac{1}{2x}$

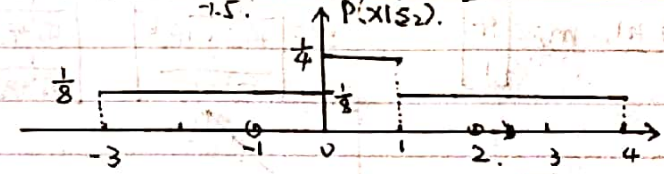
when $x \ge -1.5$.

$\hat{P}(x|S_1) = \frac{k/3}{V_3} = \frac{3/3}{2|x+3|} = \frac{1}{2(x+3)}$.



(b) $\hat{P}(x|S_2) = \frac{1}{2}[\Delta(x+1) + \Delta(x-2)]$

$= \begin{cases} \frac{1}{8} & -3 \le x < 0 \cup 1 \le x < 4 \\ \frac{1}{4} & 0 \le x < 1 \\ 0 & \text{otherwise} \end{cases}$



(c) $\hat{P}(S_1) = \frac{3}{5}$ $\qquad \hat{P}(S_2) = \frac{2}{5}$

(d) $\hat{P}(x|S_1)\hat{P}(S_1) > \hat{P}(x|S_2)\hat{P}(S_2)$. $\quad x \in S_1$

$\hat{P}(x|S_1)\hat{P}(S_1) < \hat{P}(x|S_2)\hat{P}(S_2)$, $\quad x \in S_2$

~~(e). when $x \in [-3, -1.5)$, $\frac{1}{2x} = \frac{1}{8}$ ... $x = -4$~~

~~(e) when $x \in (-\infty, -1.5)$, $\frac{1}{2x} = \frac{1}{8}$ → $x = -4$~~

(e). when $x \in (-\infty, -1.5)$, $P(x|S_1)P(S_1) = -\frac{3}{10x}$, solve for $-\frac{3}{10x} > \frac{2}{5}\cdot\frac{1}{8}$, $x > -6$.

So. when $x \in (-\infty, -1.5)$, $x \in S_1$

when $x \in [-1.5, 0)$, $P(x|S_1)P(S_1) = \frac{3}{10(x+3)}$, solve for $\frac{3}{10(x+3)} > \frac{2}{5}\cdot\frac{1}{8}$, $x < 3$.

So. when $x \in [-1.5, 0)$, $x \in S_1$

when $x \in [0, 1]$, $P(x|S_1)P(S_1) = \frac{3}{10(x+3)}$, solve $\frac{3}{10(x+3)} > \frac{2}{5}\cdot\frac{1}{4}$, $x < 0$

So when $x \in [0, 1]$, $x \in S_2$

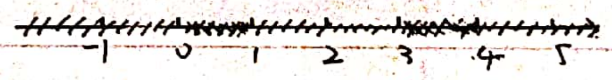
when $x \in (1, +\infty)$, $P(x|S_1)P(S_1) = \frac{3}{10(x+3)}$, solve $\frac{3}{10(x+3)} > \frac{2}{5}\cdot\frac{1}{8}$, $x < 3$

So when $x \in (1, 3)$, $x \in S_1$, when $x \in (3, 4)$, $x \in S_2$, when $x \in (4, +\infty)$, $x \in S_1$

From all above, the decision ~~boundary~~ rule is:

if $x \in (-\infty, 0) \cup (1, 3) \cup (4, +\infty)$, $\quad x \in S_1$

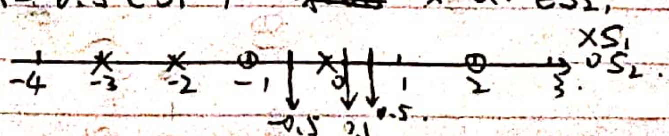
if $x \in [0, 1) \cup (3, 4)$, $\quad x \in S_2$



(f). From the plot and decision rules above,

$x = -0.5 \in S_1$, $\quad x = 0.1 \in S_2$, $\quad x = 0.5 \in S_2$

(g).



~~$x = -0.5$, the nearest 3 points, 2 belongs to S_1, 1 belongs to S_2.~~

So $x = -0.5 \in S_1$,

$x = 0.1$: $2 \in S_2$, $1 \in S_1$, so $x = 0.1 \in S_2$

$x = 0.5$: $2 \in S_2$, $1 \in S_1$, so $x = 0.5 \in S_2$

2.(a). $P(Z|\theta) = P(x_1, x_2 \cdots x_N|\theta) \overset{iid}{=\!=} \prod_{i=1}^{N} p(x_i|\theta) = \theta^N e^{-\theta \sum_{i=1}^{N} x_i}$. (for x_i drawn from θ, they all ≥ 0, so $p(x_i|\theta) = \theta e^{-\theta x_i}$.)

$$\arg\max_\theta (P(Z|\theta)) = \arg\max_\theta (\ln P(Z|\theta)) = \arg\max_\theta \left(N\ln\theta - \theta\sum_{i=1}^{N} x_i\right) = \arg\max_\theta (N\ln\theta - \theta N m)$$

$\frac{\partial}{\partial\theta}\ln(P(Z|\theta)) = \frac{N}{\theta} - Nm$, solve $\frac{N}{\theta_m} - Nm = 0$, $\hat{\theta}_{ML} = \frac{1}{m}$

(b). $\hat{\theta}_{MAP} = \arg\max_\theta \{\ln P(Z|\theta) + \ln p(\theta)\}$.

$\ln P(Z|\theta) + \ln P(\theta) = N\ln\theta - Nm\theta + \ln a - a\theta$

$\frac{\partial}{\partial\theta}(\ln P(Z|\theta) + \ln P(\theta)) = \frac{N}{\theta} - Nm - a$, $\hat{\theta}_{MAP} = \frac{N}{Nm+a}$

(c) $\hat{\theta}_{MAP} = \frac{N}{N/\hat{\theta}_{ML} + a}$

$$\lim_{\sigma_0\to\infty} \hat{\theta}_{MAP} = \lim_{\sigma_0\to\infty} \frac{N}{N/\hat{\theta}_{ML} + \frac{1}{\sigma_0}} = \frac{N}{N/\hat{\theta}_{ML}} = \hat{\theta}_{ML}$$

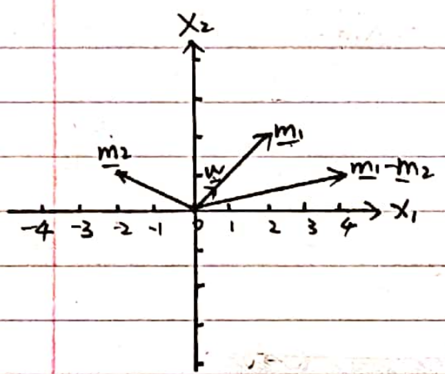
~~Because when $\sigma_0 \to \infty$, $\hat{\theta}_{MAP} \to \hat{\theta}_{ML}$, so this correspond to.~~

When $\sigma_0 \to \infty$, $a \to 0$, which means for $\forall\theta > 0$, $p(\theta) \to 0$, so the prior distribution of θ is unknown to us.

3(a). $\underline{\underline{S}}_w = \underline{\underline{S}}_1 + \underline{\underline{S}}_2 = \begin{pmatrix} \sigma_1^2+\rho_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_D^2+\rho_D^2 \end{pmatrix}$. $\underline{\underline{S}}_w^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2+\rho_1^2} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sigma_D^2+\rho_D^2} \end{pmatrix}$.

$$\underline{w} = S_w^{-1} \begin{bmatrix} m_1^{(1)} - m_1^{(2)} \\ \vdots \\ m_D^{(1)} - m_D^{(2)} \end{bmatrix} = \begin{bmatrix} (m_1^{(1)} - m_1^{(2)})/(\sigma_1^2+\rho_1^2) \\ \vdots \\ (m_D^{(1)} - m_D^{(2)})/(\sigma_D^2+\rho_D^2) \end{bmatrix}$$

(b). $\underline{m}_1 - \underline{m}_2 = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$ $\underline{w} = \begin{bmatrix} \frac{4}{4\sigma_2^2+4\rho_2^2} \\ \frac{1}{\sigma_2^2+\rho_2^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma_2^2+\rho_2^2} \\ \frac{1}{\sigma_2^2+\rho_2^2} \end{bmatrix}$ normalize $\|w\|$ to 1, $\underline{w}' = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$



(c). After projecting all data points on w, the criterion function $J(\underline{w}) = \frac{|\tilde{m}_1 - \tilde{m}_2|^2}{\tilde{S}_1^2 + \tilde{S}_2^2}$ is maximized, where $\tilde{m}_i = \underline{w}^T \underline{m}_i$, $\tilde{S}_i = w^T S_i w$, and the data points in the same classes are projected closer, the data points in different classes are projected far away from each other. $\underline{w}$ makes more sense because $m_1 - m_2$ is only a part of the criteria function, ~~they~~ it will be influenced by scatter matrices when computing the optimal projection. So $\underline{w}$ ~~is a~~ makes more sence.