



Solution 2

Prof. Krishna R. Pattipati

**Dept. of Electrical and Computer Engineering
University of Connecticut**

Contact: krishna@engr.uconn.edu (860) 486-2890

ECE 6141
Neural Networks for Classification and Optimization



■ Problem 1

$$\text{Recall } g(x) = \sigma^2 [g_1(x) - g_2(x)] = (\mu_1 - \mu_2)x - \left[\frac{1}{2}(\mu_1^2 - \mu_2^2) - \sigma^2 \ln \frac{P(\omega_1)}{P(\omega_2)} \right] > 0$$

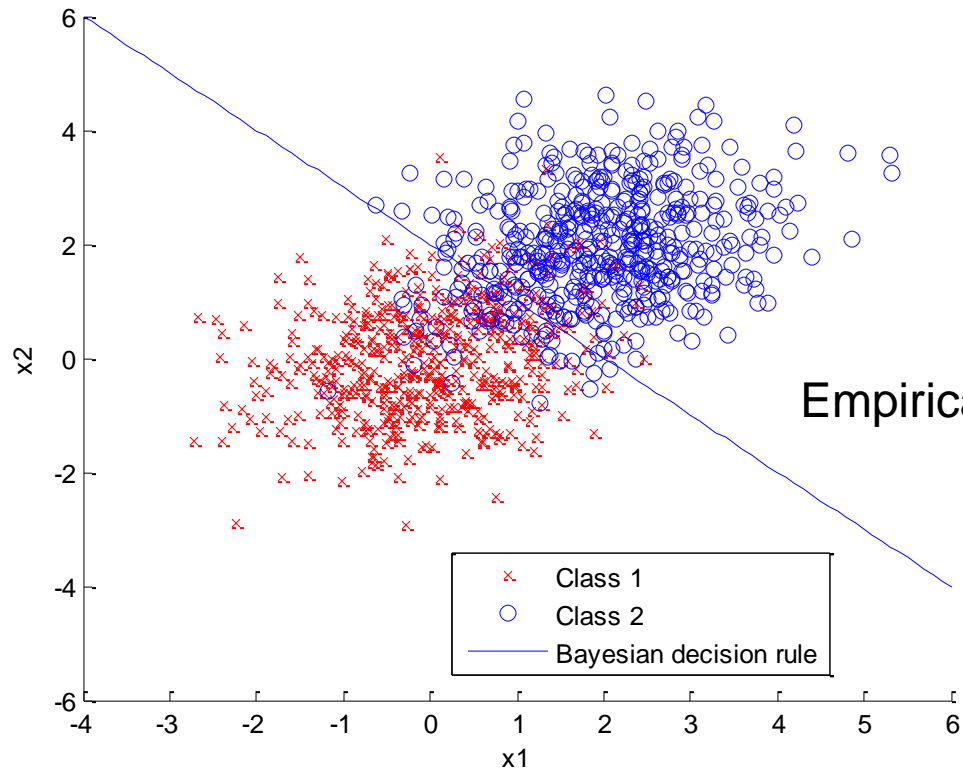
Assuming equal priors, $g(x) = x > 0$

$$P_e = \Phi\left(-\frac{d'}{2}\right) = \Phi(-10) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-10} e^{-\frac{u^2}{2}} du \approx \frac{1}{\sqrt{2\pi}10} e^{-50} \approx 7.6946 * 10^{-24}$$

$$\text{u sin g for large } x, \Phi(-x) \approx \frac{1}{\sqrt{2\pi} |x|} e^{-\frac{x^2}{2}}$$



■ Problem 2





■ Problem 2

$$\underline{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}; \underline{\mu}_2 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$\Sigma_1 = \Sigma_2 = \Sigma = \begin{bmatrix} 1 & 0.25 \\ 0.25 & 1 \end{bmatrix} \Rightarrow \Sigma^{-1} = \begin{bmatrix} 16/15 & -4/15 \\ -4/15 & 16/15 \end{bmatrix}$$

Discriminant $g(\underline{x}) = (\underline{\mu}_1 - \underline{\mu}_2)^T \Sigma^{-1} \underline{x} - \frac{1}{2} (\underline{\mu}_1^T \Sigma^{-1} \underline{\mu}_1 - \underline{\mu}_2^T \Sigma^{-1} \underline{\mu}_2)$ **assuming equal priors**

$$= -1.6x_1 - 1.6x_2 + 3.2$$

\Rightarrow class 1 if $g'(\underline{x}) = x_1 + x_2 - 2 \leq 0$; else class 2.

$$p(g(\underline{x}) | z = 1) = N(3.2, 6.4); p(g(\underline{x}) | z = 2) = N(-3.2, 6.4)$$

$$P_e = \frac{1}{2} \left[\underbrace{\int_0^{\infty} N(g; 3.2, 6.4) dg}_{\text{decision}=2} + \underbrace{\int_{-\infty}^0 N(g; -3.2, 6.4) dg}_{\text{decision}=1} \right]$$
$$= \int_{-\infty}^0 N(g; 3.2, 6.4) dg = \Phi(-1.2649) = 0.10295$$

Alternately: $p(g'(\underline{x}) | z = 1) = N(-2, 2.5); p(g'(\underline{x}) | z = 2) = N(2, 2.5)$

$$P_e = \frac{1}{2} \left[\int_0^{\infty} N(g; -2, 2.5) dg + \int_{-\infty}^0 N(g; 2, 2.5) dg \right] = 0.10295$$



■ Problem 2 (assume equal priors)

$$\Lambda = \begin{bmatrix} 0 & 1 \\ 0.005 & 0 \end{bmatrix}$$

\Rightarrow class 1 if $P(z = 2 | \underline{x}) \leq 0.005P(z = 1 | \underline{x})$; else class 2

\Rightarrow class 1 if $P(z = 1 | \underline{x}) \geq 200P(z = 2 | \underline{x})$; else class 2.

\Rightarrow class 1 if $P(\underline{x} | z = 1) \geq 200P(\underline{x} | z = 2)$; else class 2.

\Rightarrow class 1 if $\underline{\mu}_2^T \Sigma^{-1} \underline{x} \leq \frac{\underline{\mu}_2^T \Sigma^{-1} \underline{\mu}_2}{2} - \ln 200$; else class 2

\Rightarrow class 1 if $g(\underline{x}) = 1.6x_1 + 1.6x_2 + 2.0983 \leq 0$; else class 2

$$p(g(\underline{x}) | z = 1) = N(g; 2.0983, 6.4)$$

$$p(g(\underline{x}) | z = 2) = N(g; 8.4983, 6.4)$$

$$ECM = \frac{1}{2} [0.005 * \underbrace{P\{(g(\underline{x}) | z = 1) \geq 0\}}_{\text{decision}=2} + \underbrace{P\{(g(\underline{x}) | z = 2) \leq 0\}}_{\text{decision}=1}]$$

$$= \frac{1}{2} [0.7966 * 0.005 + 0.00039077] = 0.0022$$

$$\text{Empirical Risk} = \frac{1}{N} \sum_{k=1}^N \min[P(z = 2 | \underline{x}_k), 0.005P(z = 1 | \underline{x}_k)]$$



■ Problem 2 $\Lambda = \begin{bmatrix} 0 & 1000 \\ 1 & 0 \end{bmatrix}$

\Rightarrow class 1 if $1000P(z = 2 | \underline{x}) \leq P(z = 1 | \underline{x})$; else class 2

\Rightarrow class 1 if $P(z = 1 | \underline{x}) \geq 1000P(z = 2 | \underline{x})$; else class 2.

\Rightarrow class 1 if $P(\underline{x} | z = 1) \geq 1000P(\underline{x} | z = 2)$; else class 2.

\Rightarrow class 1 if $\ln P(\underline{x} | z = 1) \geq \ln 1000 + \ln P(\underline{x} | z = 2)$; else class 2.

\Rightarrow class 1 if $-\frac{1}{2}(\underline{x} - \underline{\mu}_1)^T \Sigma^{-1}(\underline{x} - \underline{\mu}_1) \geq 6.9078 - \frac{1}{2}(\underline{x} - \underline{\mu}_2)^T \Sigma^{-1}(\underline{x} - \underline{\mu}_2)$; else class 2.

\Rightarrow class 1 if $-\underline{\mu}_2^T \Sigma^{-1} \underline{x} \geq 6.9078 - \frac{\underline{\mu}_2^T \Sigma^{-1} \underline{\mu}_2}{2} = 6.9078 - 3.2 = 3.7078$; else class 2.

\Rightarrow class 1 if $-1.6x_1 - 1.6x_2 \geq 3.7078$; else class 2.

\Rightarrow class 1 if $g(\underline{x}) = x_1 + x_2 + 2.3174 \leq 0$; else class 2.

$p(g(\underline{x}) | z = 1) = N(g; 2.3174, 2.5)$; $p(g(\underline{x}) | z = 2) = N(g; 6.3174, 2.5)$

$$ECM = \frac{1}{2} \left[\underbrace{P\{(g(\underline{x}) | z = 1) > 0\}}_{\text{decision}=2} + 1000 \underbrace{P\{(g(\underline{x}) | z = 2) \leq 0\}}_{\text{decision}=1} \right]$$

$$= \frac{1}{2} [0.9286 + 1000 * 0.00003] = 0.46445$$

$$\text{Empirical Risk} = \frac{1}{N} \sum_{k=1}^N [1000 * P(\text{decision} = 1, \text{class } 2) + P(\text{decision} = 2, \text{class } 1)] = 0.463$$

Class 1	Class 2
37	0
463	500



■ Problem 2 $\Lambda = \begin{bmatrix} 0 & 1 \\ 10 & 0 \end{bmatrix}$

\Rightarrow class 1 if $P(z = 2 | \underline{x}) \leq 10P(z = 1 | \underline{x})$; else class 2

\Rightarrow class 1 if $P(z = 1 | \underline{x}) \geq 0.1P(z = 2 | \underline{x})$; else class 2.

\Rightarrow class 1 if $P(\underline{x} | z = 1) \geq 0.1P(\underline{x} | z = 2)$; else class 2.

\Rightarrow class 1 if $\ln P(\underline{x} | z = 1) \geq -2.3026 + \ln P(\underline{x} | z = 2)$; else class 2.

\Rightarrow class 1 if $-\frac{1}{2}(\underline{x} - \underline{\mu}_1)^T \Sigma^{-1}(\underline{x} - \underline{\mu}_1) \geq -2.3026 - \frac{1}{2}(\underline{x} - \underline{\mu}_2)^T \Sigma^{-1}(\underline{x} - \underline{\mu}_2)$; else class 2.

\Rightarrow class 1 if $-\underline{\mu}_2^T \Sigma^{-1} \underline{x} \geq -2.3026 - \frac{\underline{\mu}_2^T \Sigma^{-1} \underline{\mu}_2}{2} = -2.3026 - 3.2 = -5.5026$; else class 2.

\Rightarrow class 1 if $-1.6x_1 - 1.6x_2 \geq -5.5026$; else class 2.

\Rightarrow class 1 if $g(\underline{x}) = x_1 + x_2 - 3.4391 \leq 0$; else class 2.

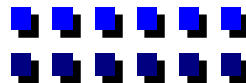
$p(g(\underline{x}) | z = 1) = N(g; -3.4391, 2.5)$; $p(g(\underline{x}) | z = 2) = N(g; 0.5609, 2.5)$

$$ECM = \frac{1}{2} [10 * \underbrace{P\{(g(\underline{x}) | z = 1) > 0\}}_{\text{decision}=2} + \underbrace{P\{(g(\underline{x}) | z = 2) \leq 0\}}_{\text{decision}=1}]$$

$$= \frac{1}{2} [10 * 0.0148 + 0.36139] = 0.2547$$

$$\text{Empirical Risk} = \frac{1}{N} \sum_{k=1}^N [P(\text{decision} = 1, \text{class } 2) + 10 * P(\text{decision} = 2, \text{class } 1)] = 0.24$$

Class 1	Class 2
492	160
8	340





■ Problem 3

Q = Quantity to buy

D = Demand with *pdf* $f(D)$ and *distribution* $F(D)$

C = Cost per item

P = Selling price

$$\text{Profit} = \begin{cases} (P - C)Q & \text{if } D > Q \\ (P - C)D & \text{if } D \leq Q \end{cases}$$

$$\text{Waste} = \begin{cases} 0 & \text{if } D > Q \\ (Q - D)C & \text{if } D \leq Q \end{cases}$$

$$\text{Net expected profit} = E_D[\pi(Q)] = \int_Q^\infty (P - C)Qf(D)dD + \int_0^Q (P - C)Df(D)dD - \int_0^Q (Q - D)Cf(D)dD$$

$$\text{Optimal } Q^* \Rightarrow \frac{dE_D[\pi(Q)]}{dQ} = 0 \Rightarrow \int_{Q^*}^\infty (P - C)f(D)dD - (P - C)Q^*f(Q^*) + (P - C)Q^*f(Q^*) - \int_0^{Q^*} Cf(D)dD = 0$$

$$\Rightarrow (P - C)[1 - F(Q^*)] - CF(Q^*) = 0 \Rightarrow (P - C) - PF(Q^*) = 0 \Rightarrow F(Q^*) = \frac{P - C}{P} \Rightarrow Q^* = F^{-1}\left(\frac{P - C}{P}\right)$$

$$\text{Second derivative: } \frac{d^2E_D[\pi(Q)]}{dQ^2} = -(P - C)f(Q^*) - Cf(Q^*) = -Pf(Q^*) < 0 \Rightarrow \text{maximum}$$



■ Problem 4

$$L(\Delta, p(\cdot)) = -\ln p(\Delta)$$

Two possibilities :

$$p^{BMA}(\Delta) = \sum_m p(\Delta | m, D) P(m | D)$$

$$p^m(\Delta) = p(\Delta | m, D)$$

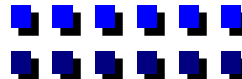
$$E_{p^{BMA}} [L(\Delta, p^{BMA})] = -\int_{\Delta} p^{BMA}(\Delta) \ln p^{BMA}(\Delta) d\Delta$$

$$E_{p^{BMA}} [L(\Delta, p^m)] = -\int_{\Delta} p^{BMA}(\Delta) \ln p^m(\Delta) d\Delta$$

$$E_{p^{BMA}} [L(\Delta, p^{BMA}) - L(\Delta, p^m)] = -\int_{\Delta} p^{BMA}(\Delta) \ln \frac{p^{BMA}(\Delta)}{p^m(\Delta)} d\Delta = -KL(p^{BMA}(\Delta) || p^m(\Delta)) \leq 0$$

$$\text{Why?} -\int_{\Delta} p^{BMA}(\Delta) \ln \frac{p^{BMA}(\Delta)}{p^m(\Delta)} d\Delta = \int_{\Delta} p^{BMA}(\Delta) \ln \frac{p^m(\Delta)}{p^{BMA}(\Delta)} d\Delta \leq \ln \int_{\Delta} p^{BMA}(\Delta) \frac{p^m(\Delta)}{p^{BMA}(\Delta)} d\Delta = \ln 1 = 0$$

Recall $\ln x$ is concave





■ Problem 5

$$a. p(x, y | \underline{\theta}) = p(x | \theta_1) p(y | x, \theta_2)$$

$$p(x | \theta_1) = \theta_1^x (1 - \theta_1)^{1-x}$$

$$p(y | x, \theta_2) = \theta_2^{[(1-x)(1-y)+xy]} (1 - \theta_2)^{[(1-x)y+x(1-y)]}$$

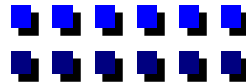
$$\therefore p(x, y | \underline{\theta}) = \theta_1^x (1 - \theta_1)^{1-x} \theta_2^{[(1-x)(1-y)+xy]} (1 - \theta_2)^{[(1-x)y+x(1-y)]}$$

In terms of a 2x2 table, $p(x, y | \underline{\theta})$ is:

	$y = 0$	$y = 1$
$x = 0$	$(1 - \theta_1)\theta_2$	$(1 - \theta_1)(1 - \theta_2)$
$x = 1$	$\theta_1(1 - \theta_2)$	$\theta_1\theta_2$

$$b. p(D | \underline{\theta}, M_2) = \prod_{i=1}^N \left[\theta_1^{x_i} (1 - \theta_1)^{1-x_i} \theta_2^{[(1-x_i)(1-y_i)+x_i y_i]} (1 - \theta_2)^{[(1-x_i)y_i+x_i(1-y_i)]} \right]; N = 7$$

Given : $\underline{x} = (1, 1, 0, 1, 1, 0, 0)$; $\underline{y} = (1, 0, 0, 0, 1, 0, 1)$





■ Problem 5

Given : $\underline{x} = (1, 1, 0, 1, 1, 0, 0)$; $\underline{y} = (1, 0, 0, 0, 1, 0, 1)$

$$\text{Let } n_1 = \sum_{i=1}^7 x_i = 4; n_0 = 7 - n_1 = 3$$

$$\text{Let } D_c = \begin{bmatrix} n_{00} & n_{01} \\ n_{10} & n_{11} \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 2 & 2 \end{bmatrix}$$

$$\begin{aligned} J_2 &= -\ln p(D | \underline{\theta}, M_2) = -[n_1 \ln \theta_1 + n_0 \ln(1 - \theta_1) + (n_{00} + n_{11}) \ln \theta_2 + (n_{01} + n_{10}) \ln(1 - \theta_2)] \\ &= -[4 \ln \theta_1 + 3 \ln(1 - \theta_1) + 4 \ln \theta_2 + 3 \ln(1 - \theta_2)] \end{aligned}$$

$$\frac{\partial J_2}{\partial \theta_1} = 0 \Rightarrow \hat{\theta}_{1,MLE} = \frac{4}{7}$$

$$\frac{\partial J_2}{\partial \theta_2} = 0 \Rightarrow \hat{\theta}_{2,MLE} = \frac{4}{7}$$

$$J_2^* = -8 \ln \frac{4}{7} - 6 \ln \frac{3}{7} = 9.5607 \Rightarrow p(D | \hat{\underline{\theta}}_{MLE}, M_2) = e^{-J_2} = 7.0443 \times 10^{-5}$$

c) Similarly, $\underline{\theta} = (\theta_{00}, \theta_{01}, \theta_{10}, \theta_{11})$

$$J_4 = -\ln p(D | \underline{\theta}, M_4) = -[2 \ln \theta_{00} + \ln \theta_{01} + 2 \ln \theta_{10} + 2 \ln(1 - \theta_{00} - \theta_{01} - \theta_{10})]$$



■ Problem 5

Given : $\underline{x} = (1, 1, 0, 1, 1, 0, 0)$; $\underline{y} = (1, 0, 0, 0, 1, 0, 1)$

$$\text{Let } n_1 = \sum_{i=1}^7 x_i = 4; n_0 = 7 - n_1 = 3$$

$$\text{Let } D_c = \begin{bmatrix} n_{00} & n_{01} \\ n_{10} & n_{11} \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 2 & 2 \end{bmatrix}$$

$$\begin{aligned} J_4 &= -\ln p(D | \underline{\theta}, M_4) = - \left[n_{00} \ln \theta_{00} + n_{01} \ln \theta_{01} + n_{10} \ln \theta_{10} + n_{11} \ln \underbrace{(1 - \theta_{00} - \theta_{01} - \theta_{10})}_{\theta_{11}} \right] \\ &= - \left[2 \ln \theta_{00} + \ln \theta_{01} + 2 \ln \theta_{10} + 2 \ln \underbrace{(1 - \theta_{00} - \theta_{01} - \theta_{10})}_{\theta_{11}} \right] \end{aligned}$$

Easy to see that $\hat{\theta}_{00} = \hat{\theta}_{10} = \hat{\theta}_{11} = \frac{2}{7}$; $\hat{\theta}_{01} = \frac{\hat{\theta}_{00}}{2} = \frac{1}{7}$

$$J_4^* = 9.4625 \Rightarrow p(D | \hat{\underline{\theta}}_{MLE}, M_4) = e^{-J_4^*} = 7.7712 \times 10^{-5}$$



Problem 5

d) Given : $\underline{x} = (1, 1, 0, 1, 1, 0, 0)$; $\underline{y} = (1, 0, 0, 0, 1, 0, 1)$

$$\text{Let } n_1 = \sum_{i=1}^7 x_i = 4; n_0 = 7 - n_1 = 3$$

$$\text{Let } D_c = \begin{bmatrix} n_{00} & n_{01} \\ n_{10} & n_{11} \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 2 & 2 \end{bmatrix}; \text{LOOCV : } D_{-1} = D_{-5} = \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix}; D_{-2} = D_{-4} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix};$$

$$D_{-3} = D_{-6} = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}; D_{-7} = \begin{bmatrix} 2 & 0 \\ 2 & 2 \end{bmatrix}$$

For model 2:

$$(i) \text{ For } D_{-1} \& D_{-5} : J = -[3 \ln \theta_1 + 3 \ln(1 - \theta_1) + 3 \ln \theta_2 + 3 \ln(1 - \theta_2)] \Rightarrow \hat{\theta}_1 = \hat{\theta}_2 = \frac{1}{2}$$

$$(ii) \text{ For } D_{-2} \& D_{-4} : J = -[3 \ln \theta_1 + 3 \ln(1 - \theta_1) + 4 \ln \theta_2 + 2 \ln(1 - \theta_2)] \Rightarrow \hat{\theta}_1 = \frac{1}{2}; \hat{\theta}_2 = \frac{2}{3}$$

$$(iii) \text{ For } D_{-3} \& D_{-6} : J = -[4 \ln \theta_1 + 2 \ln(1 - \theta_1) + 3 \ln \theta_2 + 3 \ln(1 - \theta_2)] \Rightarrow \hat{\theta}_1 = \frac{2}{3}; \hat{\theta}_2 = \frac{1}{2}$$

$$(iv) \text{ For } D_{-7} : -[4 \ln \theta_1 + 2 \ln(1 - \theta_1) + 4 \ln \theta_2 + 2 \ln(1 - \theta_2)] \Rightarrow \hat{\theta}_1 = \frac{2}{3}; \hat{\theta}_2 = \frac{2}{3}$$

$$L(2) = \ln p(D | \underline{\theta}, M_2) = 4 \ln \frac{1}{2} + 2 \left[\ln \frac{1}{2} + \ln \frac{1}{3} \right] + 2 \left[\ln \frac{1}{3} + \ln \frac{1}{2} \right] + 2 \ln \frac{1}{3} = 8 \ln \frac{1}{2} + 6 \ln \frac{1}{3} = -12.1369$$



Problem 5

d) Given : $\underline{x} = (1, 1, 0, 1, 1, 0, 0)$; $\underline{y} = (1, 0, 0, 0, 1, 0, 1)$

$$\text{Let } n_1 = \sum_{i=1}^7 x_i = 4; n_0 = 7 - n_1 = 3$$

$$\text{Let } D_c = \begin{bmatrix} n_{00} & n_{01} \\ n_{10} & n_{11} \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 2 & 2 \end{bmatrix}; \text{ LOOCV : } D_{-1} = D_{-5} = \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix}; D_{-2} = D_{-4} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix};$$

$$D_{-3} = D_{-6} = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}; D_{-7} = \begin{bmatrix} 2 & 0 \\ 2 & 2 \end{bmatrix}$$

For model 4: For $D_{-7}, \hat{\theta}_{01} = 0 \Rightarrow L(4) = -\infty$ because $p((0,1) | D_{-7}, M_4) = \ln 0 = -\infty$

So, LOOCV will pick model M_2 .

$$e) BIC(M_2) = \ln P(D | \hat{\theta}_{MLE}, M_2) - \frac{\text{dof}(M_2)}{2} \ln N = -9.5607 - \ln 7 = -11.5066$$

$$BIC(M_4) = \ln P(D | \hat{\theta}_{MLE}, M_4) - \frac{\text{dof}(M_4)}{2} \ln N = -9.4625 - \frac{3}{2} \ln 7 = -12.1814$$

$BIC(M_2) > BIC(M_4) \Rightarrow BIC$ will also pick Model 2.



Problem 6

$$\frac{p(\underline{x} | z = 1)}{p(\underline{x} | z = 2)} \geq \frac{(\lambda_{12} - \lambda_{22})P(z = 2)}{(\lambda_{21} - \lambda_{11})P(z = 1)} \Rightarrow \hat{z} = 1; \text{ otherwise } \hat{z} = 2$$

$$\Rightarrow \ln p(x | z = 1) - \ln p(x | z = 2) \geq \ln \left[\frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \right] + \ln P(z = 2) - \ln P(z = 1) \Rightarrow \text{class 1; else class 2}$$

$$\Rightarrow -\frac{1}{2} \ln(0.5) - \frac{1}{2} \frac{(x-2)^2}{0.5} + \frac{1}{2} \ln(0.2) + \frac{1}{2} \frac{(x-1.5)^2}{0.2} \geq \ln \underbrace{\left[\frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \right]}_t + \ln \frac{1}{3} - \ln \frac{2}{3} \Rightarrow \text{class 1; else class 2}$$

$$\Rightarrow 0.3466 - (x-2)^2 - 0.8047 + 2.5(x-1.5)^2 \geq \ln t - 0.6931 \Rightarrow \text{class 1; else class 2}$$

$$\Rightarrow 1.5x^2 - 3.5x + 1.7975 - \ln t \geq 0 \Rightarrow \text{class 1; else class 2}$$

$$t = 1 \Rightarrow \text{class 1 for } x \leq 0.7632 \text{ \& } x \geq 1.5701; \text{ class 2 for } x \in (0.7632, 1.5701)$$

