JHU Engineering for Professionals
Applied and Computational Mathematics
Data Mining: 625.740

Module 3 Homework Solutions

1. (a)

$$P(\omega_1|x) = \frac{p(x|\omega_1)P(\omega_1)}{p(x|\omega_1)P(\omega_1) + p(x|\omega_2)P(\omega_2)}$$
$$= \frac{1 + \left(\frac{x-a_2}{b}\right)^2}{2 + \left(\frac{x-a_1}{b}\right)^2 + \left(\frac{x-a_2}{b}\right)^2}$$
$$P\left(\omega_1 \middle| \frac{a_1+a_2}{2}\right) = \frac{1 + \left(\frac{a_1-a_2}{2b}\right)^2}{2 + 2\left(\frac{a_1-a_2}{2b}\right)^2} = \frac{1}{2} = P\left(\omega_2 \middle| \frac{a_1+a_2}{2}\right)$$

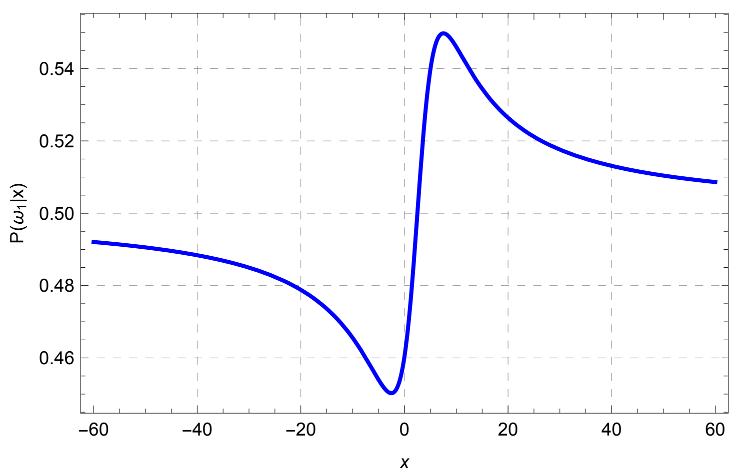

Figure 1. The graph of $P(\omega_1|x)$ vs.x for the case $a_1 = 3, a_2 = 2, b = 5$.

As $x \to -\infty, P(\omega_1|x) \to \frac{1}{2}$, and as $x \to \infty, P(\omega_1|x) \to \frac{1}{2}$

(b)

$$P(\text{error}) = P(\text{choosing } \omega_2|\omega_1)P(\omega_1) + P(\text{choosing } \omega_1|\omega_2)P(\omega_2)$$
$$= \frac{1}{2}\int_{\frac{a_1+a_2}{2}}^{\infty} p(x|\omega_1)\,dx + \frac{1}{2}\int_{-\infty}^{\frac{a_1+a_2}{2}} p(x|\omega_2)\,dx$$
$$= \frac{1}{\pi b}\int_{\frac{a_1+a_2}{2}}^{\infty} \frac{dx}{1 + \left(\frac{x-a_1}{b}\right)^2}$$
$$= \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\left|\frac{a_2-a_1}{2b}\right|.$$

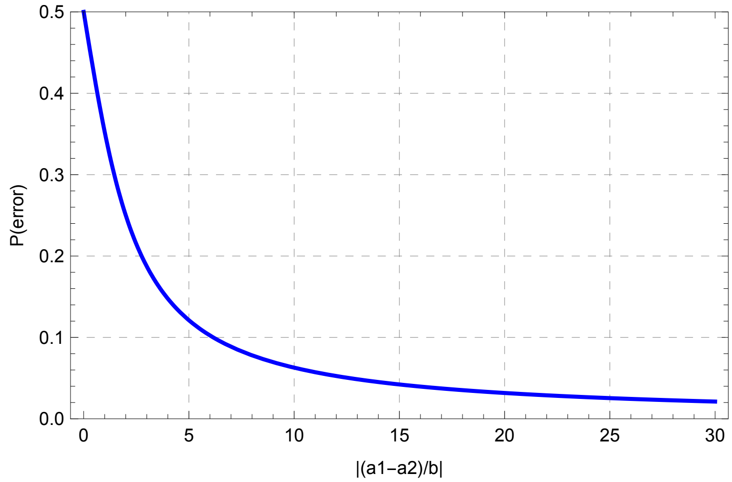

Figure 2. The graph of $P(\text{error})$ vs.$|(a_2 - a_1)/b|$.

2. (a) The mean of k is $\mathbf{E}(k) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda$.

   (b) $\mathbf{E}(k^2) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{k^2}{k!} \lambda^k = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{k}{(k-1)!} \lambda^{k-1} = \lambda e^{-\lambda} \frac{d}{d\lambda}(\lambda e^{\lambda}) = \lambda(\lambda+1)$.
   $\mathrm{Var}(k) = \mathbf{E}(k^2) - (\mathbf{E}(k))^2 = \lambda(\lambda+1) - \lambda^2 = \lambda$.

   (c) $\frac{P(k|\lambda)}{P(k-1|\lambda)} = \frac{\lambda}{k}$. If $\lambda \in$ Integers, then $\text{mode}(k) = \{\lambda, \lambda - 1\}$, otherwise $\text{mode}(k) = \lfloor \lambda \rfloor$.

   (d) Bayes decision rule is to choose C_1 if $P(k|C_1)P(C_1) > P(k|C_2)P(C_2)$. That is, choose C_1 if $k > k^* = \frac{\lambda_1 - \lambda_2}{\ln \lambda_1 - \ln \lambda_2}$.

   (e) The Bayes error rate is

$$
\begin{aligned}
P_e &= P(\text{choosing } \omega_2 | \omega_1) P(\omega_1) + P(\text{choosing } \omega_1 | \omega_2) P(\omega_2) \\
&= \frac{1}{2} \left[e^{-\lambda_1} \sum_{k=0}^{\lfloor k^* \rfloor} \frac{\lambda_1^k}{k!} + e^{-\lambda_2} \sum_{k=\lfloor k^* \rfloor + 1}^{\infty} \frac{\lambda_2^k}{k!} \right].
\end{aligned}
$$

3. We first solve the problem in one dimension, and then show that the k-dimensional solution can be reduced to one dimension by the symmetry of the problem.

   I. Consider the one-dimensional problem $p(\mathbf{x}|\omega_i) \sim N(\mu_i, \sigma^2)$ for two-categories with $P(\omega_1) = P(\omega_2) = \frac{1}{2}$.

$$
\begin{aligned}
P_e &= P(\text{choosing } \omega_2|\omega_1)P(\omega_1) + P(\text{choosing } \omega_1|\omega_2)P(\omega_2) \\
&= \frac{1}{2}\int_{\frac{\mu_1+\mu_2}{2}}^{\infty} p(x|\omega_1)\,dx + \frac{1}{2}\int_{-\infty}^{\frac{\mu_1+\mu_2}{2}} p(x|\omega_2)\,dx \\
&= \frac{1}{2\sigma\sqrt{2\pi}}\int_{\frac{\mu_1+\mu_2}{2}}^{\infty} \exp\left\{-\frac{(x-\mu_1)^2}{2\sigma^2}\right\}dx + \frac{1}{2\sigma\sqrt{2\pi}}\int_{-\infty}^{\frac{\mu_1+\mu_2}{2}} \exp\left\{-\frac{(x-\mu_2)^2}{2\sigma^2}\right\}dx \\
&= \frac{1}{\sigma\sqrt{2\pi}}\int_{\frac{\mu_1+\mu_2}{2}}^{\infty} \exp\left\{-\frac{(x-\mu_1)^2}{2\sigma^2}\right\}dx \\
&= \frac{1}{\sqrt{2\pi}}\int_{\frac{|\mu_1-\mu_2|}{2\sigma}}^{\infty} \exp\left\{-\frac{u^2}{2}\right\}du.
\end{aligned}
$$

   II. Now, $\frac{1}{\sqrt{2\pi}}\int_{\frac{|\mu_1-\mu_2|}{2\sigma}}^{\infty} \exp\left\{-\frac{u^2}{2}\right\}du = -\frac{1}{\sqrt{2\pi}}\int_a^{\infty} \frac{1}{u}\frac{d}{du}\exp\left\{-\frac{u^2}{2}\right\}du$
   $= \frac{1}{a\sqrt{2\pi}}\exp\left\{-\frac{a^2}{2}\right\} - \int_a^{\infty}\frac{\exp\left\{-\frac{u^2}{2}\right\}}{u^2}du < \frac{1}{a\sqrt{2\pi}}\exp\left\{-\frac{a^2}{2}\right\}$
   and thus $P_e \to 0$ as $a = \frac{|\mu_1-\mu_2|}{2\sigma} \to \infty$.

   III. Consider the k-dimensional case. Let $\mathbf{x}' = \mathbf{x} - \mu_{\mathbf{1}}$. Let $\mathbf{x}'' = P\mathbf{x}'$, where P is a rotation that takes $\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1$ to be along the x_1-axis. The rotation matrix P can be found by the Gram-Schmidt process.

$$
\begin{aligned}
P_e &= \frac{1}{(2\pi)^{\frac{d}{2}}}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\int_{\frac{|\boldsymbol{\mu}_1-\boldsymbol{\mu}_2|}{2\sigma}}^{\infty} \exp\left\{-\frac{x_1^2+x_2^2\cdots+x_k^2}{2}\right\}dx_1dx_2\cdots dx_k \\
&= \frac{1}{\sqrt{2\pi}}\int_{\frac{|\boldsymbol{\mu}_1-\boldsymbol{\mu}_2|}{2\sigma}}^{\infty} \exp\left\{-\frac{u^2}{2}\right\}du
\end{aligned}
$$

   IV. $P_e < \frac{1}{a\sqrt{2\pi}}\exp\left\{-\frac{a^2}{2}\right\}$. As $k \to \infty, a = \frac{|\boldsymbol{\mu}_1-\boldsymbol{\mu}_2|}{2\sigma} = \frac{\sqrt{\sum_{k=1}^{\infty} m_k^2}}{2\sigma} \to \infty$ and $P_e \to 0$.

4. Recall that the minimum risk decision rule is to decide ω_1 if

$$\frac{P(\mathbf{x}|\omega_1)}{P(\mathbf{x}|\omega_2)} > \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \frac{P(\omega_2)}{P(\omega_1)}.$$

We can write this inequality in terms of a discriminant function

$$g(\mathbf{x}) = \log \frac{P(\mathbf{x}|\omega_1)}{P(\mathbf{x}|\omega_2)} + \log \frac{P(\omega_1)}{P(\omega_2)} + \log \frac{\lambda_{21} - \lambda_{11}}{\lambda_{12} - \lambda_{22}} > 0$$

The likelihood ratio for a classifier with independent binary features is

$$\frac{P(\mathbf{x}|\omega_1)}{P(\mathbf{x}|\omega_2)} = \prod_{i=1}^{d} \left(\frac{p_i}{q_i}\right)^{x_i} \left(\frac{1-p_i}{1-q_i}\right)^{1-x_i}$$

so that

$$g(\mathbf{x}) = \sum_{i=1}^{d} \left[x_i \log \frac{p_i}{q_i} + (1-x_i) \log \frac{1-p_i}{1-q_i}\right] + \log \frac{P(\omega_1)}{P(\omega_2)} + \log \frac{\lambda_{21} - \lambda_{11}}{\lambda_{12} - \lambda_{22}}$$

or

$$g(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0,$$

where

$$\begin{aligned} w_i &= \log \frac{p_i}{q_i} - \log \frac{1-p_i}{1-q_i}, \quad i = 1, \ldots, d \\ w_0 &= \sum_{i=1}^{d} \log \frac{1-p_i}{1-q_i} + \log \frac{P(\omega_1)}{P(\omega_2)} + \log \frac{\lambda_{21} - \lambda_{11}}{\lambda_{12} - \lambda_{22}}. \end{aligned}$$