ComS 573 Machine Learning

Problem Set 1

Note: Please do not hesitate to contact the instructor or the TA if you have difficulty understanding or getting started with solving any of the problems.

1. (20 pts.) Consider a two category classification problem with one dimensional feature x. Assume that the priors are $P(\omega_1) = 1/3$ and $P(\omega_2) = 2/3$, and that the class-conditional distributions have normal densities $p(x|\omega_1) \sim N(1,1)$ and $p(x|\omega_2) \sim N(3,1)$, where

$$N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

   (a) Derive the Bayes decision rule for minimum-error-rate classification.

   (b) Let $\lambda_{ij} = \lambda(\alpha_i|\omega_j)$ be the loss incurred for deciding ω_i when the true category is ω_j. Assume $\lambda_{11} = 0$, $\lambda_{12} = 2$, $\lambda_{21} = 1$, $\lambda_{22} = 0$. Derive the Bayes decision rule for the minimum risk classification.

2. (20 pts.) Consider the Binary Independence Model for text document. Given a vocabulary $V : (w_1, \ldots, w_d)$ of all English words (and tokens), assume that a text document is represented as a vector of binary features $\vec{x} = (x_1, \ldots, x_d)^t$ such that x_i is 1 if the word w_i appears in the document, and x_i is 0 otherwise. We want to classify text documents into c categories. Let $P(\omega_j)$ be the prior probability for the class ω_j for $j = 1, \ldots, c$. Assume that the components of $\vec{x}$ are statistically independent given the category (Naive Bayes model), i.e.,

$$P(\vec{x}|\omega_j) = \prod_{i=1}^{d} P(x_i|\omega_j).$$

   Assume we could estimate the following probability from the training data

$$p_{ij} = P(x_i = 1|\omega_j) \quad i = 1, \ldots, d, \quad j = 1, \ldots, c.$$

   Show that the minimum probability of error is achieved by the following decision rule: Decide ω_k if $g_k(\vec{x}) \geq g_j(\vec{x})$ for all $j \neq k$, where the discriminant function is given in the form of

$$g_j(\vec{x}) = \sum_{i=1}^{d} c_{ij} x_i + b_j.$$

   Give the expressions of c_{ij} and b_j.

ComS-573 HW 1

Kanak Choudhury

Problem 1:

a)

We have, $P(\omega_1) = 1/3$ & $P(\omega_2) = 2/3$

and, $P(x|\omega_1) \sim N(1,1)$ & $P(x|\omega_2) \sim N(3,1)$

Posterior dist^n, $P(\omega_i|\underset{\sim}{x}) = \dfrac{P(\underset{\sim}{x}|\omega_i)P(\omega_i)}{P(\underset{\sim}{x})}$ $\quad i=1,2$

$\Rightarrow P(\omega_i|\underset{\sim}{x}) \propto P(\underset{\sim}{x}|\omega_i)P(\omega_i)$

Bayes decision rule can be written as –

$$g(x) = \begin{cases} P(\omega_1|\underset{\sim}{x}) > P(\omega_2|\underset{\sim}{x}) & \Rightarrow \omega_1 \text{ class} \\ \text{Otherwise} & \Rightarrow \omega_2 \text{ class.} \end{cases}$$

So, $P(\omega_1|\underset{\sim}{x}) > P(\omega_2|\underset{\sim}{x})$

$\Rightarrow P(\underset{\sim}{x}|\omega_1)P(\omega_1) > P(\underset{\sim}{x}|\omega_2)P(\omega_2)$

$\Rightarrow P(\underset{\sim}{x}|\omega_1) > 2\, P(\underset{\sim}{x}|\omega_2)$

$\Rightarrow \cancel{\frac{1}{\sqrt{2\pi}}} \exp\left[-\frac{1}{2}(x-1)^2\right] > 2 \cancel{\frac{1}{\sqrt{2\pi}}} \exp\left[-\frac{1}{2}(x-3)^2\right]$

$\Rightarrow -\frac{1}{2}(x-1)^2 > \ln 2 - \frac{1}{2}(x-3)^2$

$\Rightarrow (x-3)^2 - (x-1)^2 > \ln 4$

$\Rightarrow \cancel{x^2} - 6x + 9 - \cancel{x^2} + 2x - 1 > \ln 4$

$\Rightarrow -4x + 8 > \ln 4$

$\Rightarrow x < -\frac{\ln 2}{2} + 2$

So, $x < 2 - \frac{\ln 2}{2}$

Thus, Bayes decision rule is,

$$g(x) = \begin{cases} \text{if } x < 2 - \frac{\ln 2}{2} & \Rightarrow \omega_1 \text{ class} \\ \text{otherwise} & \Rightarrow \omega_2 \text{ class.} \end{cases}$$

b) We have, $\lambda_{11} = 0$, $\lambda_{12} = 2$, $\lambda_{21} = 1$ & $\lambda_{22} = 0$

Risk function can be written as—

$$R(\alpha_1 | X) = \lambda_{11} P(\omega_1 | X) + \lambda_{12} P(\omega_2 | X) \quad \text{— (I)}$$

and $$R(\alpha_2 / X) = \lambda_{21} P(\omega_1 / X) + \lambda_{22} P(\omega_2 / X) \quad \text{— (II)}$$

From (I) $\Rightarrow$

$$R(\alpha_1 | X) = 2 P(\omega_2 / X)$$

and from (II) $R(\alpha_2 / X) = P(\omega_1 / X)$

Bayes decision rule based on minimum risk classification,

$$g(x) = \begin{cases} R(\alpha_1 | X) < R(\alpha_2 | X) & \Rightarrow \omega_1 \text{ class} \\ \text{otherwise} & \Rightarrow \omega_2 \text{ class.} \end{cases}$$

So, $R(\alpha_1 | X) < R(\alpha_2 | X)$

$$\Rightarrow 2 P(\omega_2 | X) < P(\omega_1 | X)$$

$$\Rightarrow 2 \cdot 2 \cdot \cancel{\frac{1}{\sqrt{2\pi}}} \exp\left[-\frac{1}{2}(x-3)^2\right] < \cancel{\frac{1}{\sqrt{2\pi}}} \exp\left[-\frac{1}{2}(x-1)^2\right]$$

$$\Rightarrow -4x + 8 > 4 \ln 2$$

$$\Rightarrow x < 2 - \ln(2)$$

Thus, Bayes decision rule for the minimum risk classification.

$$g(x) = \begin{cases} \text{if } x < 2 - \ln(2) & \Rightarrow \omega_1 \text{ class.} \\ \text{otherwise} & \Rightarrow \omega_2 \text{ class.} \end{cases}$$

Problem 2:

Based on the given information, we can write the decision rule as —

$$P(\omega_k|x) \geq P(\omega_j|x) \qquad \forall\, j \neq k.$$

$$\Rightarrow P(\underset{\sim}{x}|\omega_k)\, P(\omega_k) \geq P(\underset{\sim}{x}|\omega_j)\, P(\omega_j)$$

$$\Rightarrow \left[\prod_{i=1}^{d} P_{ik}^{x_i}(1-P_{ik})^{1-x_i}\right] P(\omega_k) \geq \left[\prod_{i=1}^{d} P_{ij}^{x_i}(1-P_{ij})^{1-x_i}\right] P(\omega_j)$$

~~Since $x_i|\omega_j$ Bernoulli (P~~

Since, $x_i|\omega_j \overset{ind}{\sim} \text{Bernoulli}(P_{ij})$ ~~and for~~ $\forall\, i$

$P(\underset{\sim}{x}|\omega_k) = \prod_{i=1}^{d} P(x_i|\omega_j)$ since independent

$$\Rightarrow \sum_{i=1}^{d}\left[x_i \ln P_{ik} + (1-x_i)\ln(1-P_{ik})\right] + \ln P(\omega_k) \geq$$

$$\sum_{i=1}^{d}\left[x_i \ln P_{ij} + (1-x_i)\ln(1-P_{ij})\right] + \ln P(\omega_j)$$

$$\Rightarrow \sum_{i=1}^{d} x_i \ln P_{ik} + \sum_{i=1}^{d}\ln(1-P_{ik}) - \sum_{i=1}^{d} x_i \ln(1-P_{ik}) + \ln P(\omega_k) \geq$$

$$\sum_{i=1}^{d} x_i \ln P_{ij} + \sum_{i=1}^{d}\ln(1-P_{ij}) - \sum_{i=1}^{d} x_i \ln(1-P_{ij}) + \ln P(\omega_j)$$

$$\Rightarrow \sum_{i=1}^{d} x_i \ln \frac{P_{ik}}{1-P_{ik}} + \sum_{i=1}^{d} \ln(1-P_{ik}) + \ln P(\omega_k) \geq$$
$$\sum_{i=1}^{d} x_i \ln \frac{P_{ij}}{1-P_{ij}} + \sum_{i=1}^{d} \ln(1-P_{ij}) + \ln P(\omega_j)$$

Let, $c_{ij} = \ln\left(\frac{P_{ij}}{1-P_{ij}}\right)$ —— ①

and $b_j = \sum_{i=1}^{d} \ln(1-P_{ij}) + \ln P(\omega_j)$ —— ②

So, $\sum_{i=1}^{d} x_i c_{ik} + b_k \geq \sum_{i=1}^{d} x_i c_{ij} + b_j$

$\Rightarrow g_k(\underset{\sim}{x}) \geq g_j(\underset{\sim}{x})$

So, $g_j(\underset{\sim}{x}) = \sum c_{ij} x_i + b_j$

where c_{ij} and b_j are given in eq ① & ② respectively.

Thus, $c_{ij} = \ln\left(\frac{P_{ij}}{1-P_{ij}}\right)$

and $b_j = \sum_{i=1}^{d} \ln(1-P_{ij}) + \ln P(\omega_j)$