Introduction to Machine Learning - HW2 - Q3

Professors: Abolghasemi & Arabi
Student: Mohamad Mahdi Samadi
Student ID: 810101465

For a binary classifier with a single input x and equal prior probabilities, below conditional probabilities are given:

$$P(x|\bar{y}) = N(0,\ \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-1}{2}\left(\frac{x}{\sigma}\right)^2}$$

$$P(x|y) = N(2,\ \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-1}{2}\left(\frac{x-2}{\sigma}\right)^2}$$

The cost matrix is:

$$\begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix} = \begin{pmatrix} 0 & a \\ a^2 & 0 \end{pmatrix}$$

$\lambda_{i,j}$ is the cost we have to pay if the true class of a sample is i but was predicted as j. Obviously $\lambda_{i,i}$ is zero and λ is defined for off-diagonal entries.

$$R(y_j|x) = \sum_{i=1}^{N} \lambda_{i,j}P(y_i|x)$$

$$R(y_1|x) = \lambda_{1,1}P(y_1|x) + \lambda_{2,1}P(y_2|x) = \alpha\, P(y_2|x)$$

$$R(y_2|x) = \lambda_{1,2}P(y_1|x) + \lambda_{2,2}P(y_2|x) = \alpha^2 P(y_1|x)$$

The decision threshold is where two risk functions cross each other. Assuming α is a non-zero factor:

$$R(y_1|x) = R(y_2|x) \rightarrow \alpha\, P(y_2|x) = \alpha^2 P(y_1|x) \rightarrow P(y_2|x) = \alpha\, P(y_1|x)$$

Using Bayes law we have:

$$\frac{P(x|y_2)P(y_2)}{P(x)} = \alpha\, \frac{P(x|y_1)P(y_1)}{P(x)}$$

Knowing that $P(x)$ is the same and prior probabilities are equal we have:

$\alpha\, P(x|y_2) = P(x|y_1)$

$$e^{\frac{-1}{2}(\frac{x-2}{\sigma})^2} = \alpha\, e^{\frac{-1}{2}(\frac{x}{\sigma})^2} \rightarrow \alpha = e^{\frac{1}{2}[(\frac{x}{\sigma})^2-(\frac{x-2}{\sigma})^2]} = e^{\frac{1}{2}[\frac{2x-2}{\sigma}\times\frac{2}{\sigma}]} = e^{2\frac{x-1}{\sigma^2}}$$

Taking a logarithm leads us to:

$$ln(\alpha) = \frac{2x-2}{\sigma^2} \rightarrow 2x - 2 = \sigma^2 ln(\alpha) \rightarrow x = 1 + \frac{\sigma^2}{2} ln(\alpha)$$

For $\alpha < 1$, $\alpha > \alpha^2$, so the cost of mispredicting actual class $y = 0$ is more than the other one.

For $\alpha > 1$, $\alpha < \alpha^2$, so the cost of mispredicting actual class $y = 1$ is more than the other one.

A greater α (assuming it means α>1) tends to mispredict actual class $y = 0$ as little as possible, leads to a closer decision boundary to the $y = 1$ class. So greater α is, the more precise the model would be for class $y = 1$. As it predicts $\hat{y} = 1$ when it is really confident about it and predicts $\hat{y} = 0$ a lot.