1 Homework Problem

Question 1

Consider the hypothetical problem of designing a classifier to separate two kinds of fish: salmon and sea bass. Suppose a fisherman wants to predict what fish he will catch next. Assume that there is no other type of fish and that the number of sea bass is **twice** as that of salmons. The number of salmons depends on the month of the year (X_1) and the water temperature (X_2). For simplicity, you can assume that the input variables X_1, X_2 have been quantized (i.e. are discrete). Specifically, the likelihood of catching a salmon follows a bivariate normal distribution $\boldsymbol{N}(\boldsymbol{\mu}, \mathbf{I})$, where $\boldsymbol{\mu} = [\text{March}, 16^o\text{Celcius}]$, and $\mathbf{I}$ is the identity matrix. The likelihood of finding a sea bass is uniform between January and April, and zero during the remaining 8 months (from May to December) and does not depend on the water temperature.

Using the Bayes rule predict what is the most probable fish to catch at 16^o Celcius during March.

Solution: Let's denote C the type of fish, with $C = c_1$ for salmon and $C = c_2$ for sea bass and $\boldsymbol{x} = [X_1, X_2]$ the 2-dimensional input variable. We are asked to predict the fish type when $\boldsymbol{x} = \boldsymbol{\mu}$ given the following facts:

- The prior probability of sea bass is twice as that of salmons, thus $P(C = c_2) = 2P(C = c_1)$.
- The class-conditional probability density for c_1 can be written as
$$p(\boldsymbol{x}|C = c_1) = \frac{1}{\sqrt{(2\pi)^2|\mathbf{I}|}} exp(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T |\mathbf{I}|^{-1} (\boldsymbol{x} - \boldsymbol{\mu})) \tag{1}$$
- Since sea bass appear only during 4 months and with equal probability within each month, the class-conditional probability density for c_2 can be formulated as
$$p(\boldsymbol{x}|C = c_2) = \begin{cases} 1/4 & \text{if } X_1 \text{ is between January and April} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

The posterior probability of each class can be calculated using the Bayes rule and the Maximum a posteriori (MAP) estimation criterion:

$$f(\boldsymbol{x}) = \arg\max_{k=1,2} P(c_k|\boldsymbol{x}) = \arg\max_{k=1,2} \frac{P(c_k)p(\boldsymbol{x}|c_k)}{p(\boldsymbol{x})} = \arg\max_{k=1,2} P(c_k)p(\boldsymbol{x}|c_k) \tag{3}$$

where the evidence factor (denominator) was ignored because it does not depend on the class C.

1. For salmon : $P(C = c_1)p(\boldsymbol{x} = \boldsymbol{\mu}|C = c_1) = P(C = c_1)\frac{1}{\sqrt{(2\pi)^2 \cdot 1}}exp(0) = 0.16P(C = c_1)$
2. For sea bass: $P(C = c_2)p(\boldsymbol{x} = \boldsymbol{\mu}|C = c_2) = 2P(C = c_1) \cdot (1/4) = 0.5P(C = c_1)$

The probability to catch a sea bass is much higher therefore the classifier will return $C = c_2$.

Question 2

Assuming 0/1 Loss (in which $\lambda = 1$ for misclassifications or 0 otherwise), calculate the overall Bayes risk for $\boldsymbol{x} = \boldsymbol{\mu}$.

Solution:

$$R(\boldsymbol{x}) = \sum_{k=1}^{2} R(\alpha_k|\boldsymbol{x}) = \sum_{k=1}^{2}(1 - P(C_k|\boldsymbol{x})) = 1 - P(C_1|\boldsymbol{x}) + 1 - P(C_2|\boldsymbol{x}) \tag{4}$$

For $\boldsymbol{x} = \boldsymbol{\mu}$ from the previous derivation:

$$R(\boldsymbol{\mu}) = 1 - 0.16P(C_1) + 1 - 0.5P(C_1) = 2 - 0.66P(C_1)$$

If there exists no other type of fish, then $P(C_1) + P(C_2) = 1$, thus $P(C_1) = 0.33$, which results to $R(\boldsymbol{\mu}) = 1.78$.