Tutorial 4 : MAP

1 Theory

Statistical approach to classification, assuming input and output are random variables with some known probability distribution.

Notation

Ω - Finite set of classes $\omega_i \in \Omega,\ i = 1, 2, ..., N$.

X - Input space $x \in X$.

We assume that the following distributions are known:

- $\{p(\omega_1), p(\omega_2), ..., p(\omega_n)\}$ - A posteriori probabilities the classes.
- $\{p(x|\omega_1), p(x|\omega_2), ..., p(x|\omega_n)\}$ - Probability of an input x given a class ω_i (if X is a continuous space, then $p(x|\omega_i)$ is a probability density).

Maximum A Posteriori (MAP)

$$\hat{\omega}_{\text{MAP}} \triangleq \arg\max_{\omega} p(\omega|x).$$

Recall Bayes' rule:

$$p(\omega|x) = \frac{p(x|\omega)p(\omega)}{p(x)}.$$

Using the latter, we can rewrite our MAP estimator as

$$\hat{\omega}_{\text{MAP}} \triangleq \arg\max_{\omega} p(\omega|x) = \arg\max_{\omega} \frac{p(x|\omega)p(\omega)}{p(x)} = \arg\max_{\omega} p(x|\omega)p(\omega).$$

$$\Rightarrow \boxed{\hat{\omega}_{\text{MAP}} = \arg\max_{\omega} p(x|\omega)p(\omega)}$$

As shown in the lecture, this classifier minimizes both the conditional error probability, and the average error probability.

2 Practice

Question 1

A transmitted message S is either 1 or 0 is with probability p and $1-p$ respectively. Unfortunately the message gets corrupted by a Gaussian noise N with zero mean ($\mu = 0$) and unit variance $\sigma^2 = 1$. Hence, the received message Y is given by

$$Y = S + N.$$

Compute the MAP estimate of S given $Y = y$.

Solution

The MAP estimator is given by

$$\begin{aligned}
\hat{S}_{\text{MAP}} &= \underset{s=\{0,1\}}{\arg\max}\ P(S = s|Y = y) \\
&= \arg\max \left\{ f_{Y|S}(y|0)P(S = 0),\ f_{Y|S}(y|1)P(S = 1) \right\} \\
&= \arg\max \left\{ f_{Y|S}(y|0)(1-p),\ f_{Y|S}(y|1)p \right\}
\end{aligned}$$

Notice that given $S = 0$, Y becomes equal to the noise N, and therefore

$$f_{Y|S}(y|0) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}.$$

Given that $S = 1$, Y becomes $Y = N + 1$, hence

$$f_{Y|S}(y|1) = \frac{1}{\sqrt{2\pi}} e^{-(y-1)^2/2}.$$

Consequently, $\hat{S}_{\text{MAP}} = 1$ if

$$\begin{aligned}
&\frac{1}{\sqrt{2\pi}} e^{-(y-1)^2/2} p \geq \frac{1}{\sqrt{2\pi}} e^{-y^2/2}(1-p) \\
\Rightarrow &-\frac{(y-1)^2}{2} + \log(p) \geq -\frac{y^2}{2} + \log(1-p) \\
\Rightarrow &\, y \geq \frac{1}{2} + \log\left(\frac{1-p}{p}\right).
\end{aligned}$$

Hence,

$$\hat{S}_{\text{MAP}}(y) = \begin{cases} 1, & y \geq \frac{1}{2} + \log\left(\frac{1-p}{p}\right), \\ 0, & \text{o.w.} \end{cases}$$

As an example, for $p = 1 - p = \frac{1}{2}$ we get that $S = 1$ if $y \geq \frac{1}{2}$, i.e., the threshold is right in the middle of 0 and 1.

Question 2

Let X be a continuous random variable with the following PDF:

$$f_X(x) = \begin{cases} 2x, & 0 \leq x \leq 1, \\ 0, & \text{o.w.} \end{cases}$$

In addition, it is given that $Y|X = x \sim \textit{Geometric}(x)$.

Given $Y = 3$, find the MAP estimate of X.

Remainder: The geometric distribution gives the probability that the first occurrence of success requires k independent trials, each with success probability p:

$$P(Y = k) = p(1-p)^{k-1}.$$

Solution

We know that $Y|X = x \sim Geometric(x)$, so

$$P_{Y|X}(y|x) = x(1-x)^{y-1}.$$

Therefore,

$$P_{Y|X}(3|x) = x(1-x)^2.$$

The MAP estimate of X is given by

$$\hat{X}_{\text{MAP}} = \arg\max_{x\in[0,1]} P_{Y|X}(3|x) f_X(x) = x(1-x)^2 \cdot 2x = \arg\max_{x\in[0,1]} 2x^2(1-x)^2.$$

Setting the derivative to 0 we get

$$4x(1-x)^2 - 4x^2(1-x) = 0$$

Solving for x (and checking for maximization criteria), we obtain the MAP estimate as

$$\hat{X}_{\text{MAP}} = \frac{1}{2}.$$

Question 3

It is given that $X = \mathbb{R},^2$ and that $\Omega = \{\omega_1, \omega_2, \omega_3\}$.
The class prior probabilities are uniform: $p(\omega_1) = p(\omega_2) = p(\omega_3) = \frac{1}{3}$.
The class conditional probability is Gaussian, $p(x|\omega_i) \sim \mathcal{N}(\mu_i, \Sigma)$, where

$$\mu_1 = \begin{bmatrix} 0 \\ 3 \end{bmatrix}, \ \mu_2 = \begin{bmatrix} 3 \\ 0 \end{bmatrix}, \ \mu_3 = \begin{bmatrix} -3 \\ 0 \end{bmatrix}, \ \Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}.$$

(a) What is the Bayes optimal decision rule? What are the decision boundaries in the plane $\mathbb{R}^2$?

(b) Does the decision boundary for the Gaussian case always have the same shape? What it depends on?

Solution

(a) We need to compute the conditions for which $P(\omega_i|x) \geq P(\omega_j|x)$ $(i, j = 1, 2, 3, \ i \neq j)$:

$$\begin{aligned}
& P(\omega_i|x) \geq P(\omega_j|x) \\
\Leftrightarrow\ & P(x|\omega_i)p(\omega_i) \geq P(x|\omega_j)p(\omega_j) \\
\Leftrightarrow\ & P(x|\omega_i) \geq P(x|\omega_j) \\
\Leftrightarrow\ & \frac{1}{\sqrt{4\pi^2|\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu_i)^T\Sigma^{-1}(x-\mu_i)\right) \geq \frac{1}{\sqrt{4\pi^2|\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu_j)^T\Sigma^{-1}(x-\mu_j)\right) \\
\Leftrightarrow\ & \frac{1}{2}(x-\mu_i)^T\Sigma^{-1}(x-\mu_i) \leq \frac{1}{2}(x-\mu_j)^T\Sigma^{-1}(x-\mu_j) \\
\Leftrightarrow\ & ||x-\mu_i||_2^2 \leq ||x-\mu_j||_2^2 \\
\Leftrightarrow\ & ||x-\mu_i||_2 \leq ||x-\mu_j||_2
\end{aligned}$$

Thus, the decision boundary is a line in the plane $\mathbb{R}^2$.

(b) Consider the following inequality

$$\frac{1}{2}(x-\mu_i)^T\Sigma^{-1}(x-\mu_i) \leq \frac{1}{2}(x-\mu_j)^T\Sigma^{-1}(x-\mu_j).$$

Therefore, in general, the decision boundaries are a quadratic polynomial and their shape depends on Σ_i. Thus, in $\mathbb{R}^2$ the decision boundaries can be a circle, ellipse, parabola, hyperbola or a line (as in our case since all vectors have the same Σ).