Statistical learning theory

Joris LIMONIER

September 24, 2022

Contents

1 Proposed exercise lecture 02

1.1 Home Exercise

Demonstrate that the MAP (Maximum A Posteriori) classification rule is quadratic with respect to x^*. We have the following:

$$\hat{y}^* = \arg \max_{k=1,\ldots,k} \mathbb{P}\left(Y = k | X = x^*\right) \tag{1}$$

We compute the probability:

$$\begin{aligned}
\arg \max_{k=1,\ldots,k} \mathbb{P}\left(Y = k | X = x^*\right) \underbrace{=}_{\text{Bayes}} & \frac{\mathbb{P}(Y = k)\mathbb{P}(X = x^* | Y = k)}{\mathbb{P}(X = x^*)} \\
& \propto \mathbb{P}(Y = k)\mathbb{P}(X = x^* | Y = k) \\
& \propto \hat{\Pi}_k \mathcal{N}\left(x^*, \hat{\mu}_k, \hat{\Sigma}_k\right)
\end{aligned}$$

$$
\begin{aligned}
\hat{y}^* &= \arg\max_{k=1,\dots,k} \hat{\Pi}_k \mathcal{N}\left(x^*, \hat{\mu}_k, \hat{\Sigma}_k\right) \\
&= \arg\max_{k=1,\dots,k} \log\left[\hat{\Pi}_k\right] + \log\left[\mathcal{N}\left(x^*, \hat{\mu}_k, \hat{\Sigma}_k\right)\right] \\
&= \arg\max_{k=1,\dots,k} \log\left[\hat{\Pi}_k\right] + \log\left[\frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x^*-\mu)^\top \Sigma^{-1}(x^*-\mu)}\right] \\
&= \arg\max_{k=1,\dots,k} \quad \underbrace{\log\left[\hat{\Pi}_k\right]}_{\text{Constant w.r.t. } x^*} \quad + \underbrace{\left[-\frac{1}{2}(x^*-\mu)^\top \Sigma^{-1}(x^*-\mu)\right]}_{\text{Quadratic in } x^*} \\
&- \underbrace{\log\left[(2\pi)^{N/2}|\Sigma|^{1/2}\right]}_{\text{Constant w.r.t. } x^*}
\end{aligned}
$$

The second term is Quadratic in x^*, the others are constant w.r.t. x^*, which completes the proof.