Statistical learning theory

Joris LIMONIER

November 29, 2022

Contents

1 Proposed exercise lecture 02 **1**
1.1 Home Exercise 1

2 Proposed exercise lecture 05 **2**
2.1 In-class Exercise 2

3 Proposed exercise lecture 09 **3**
3.1 In-class Exercise 3

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:

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

We compute the probability:

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

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

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

2 Proposed exercise lecture 05

2.1 In-class Exercise

Compute the derivative with respect to the means.

$$
\begin{aligned}
\nabla_{\mu_k} \mathcal{L}(\theta) &= \nabla_{\mu_k} \sum_{n=1}^{N} \log\left[\sum_{k=1}^{K} \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)\right] \\
&= \sum_{n=1}^{N} \frac{\pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)}{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)} \\
&= \sum_{n=1}^{N} \frac{\pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)}{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)} \Sigma^{-1}(x_n - \mu_k)
\end{aligned}
$$

3 Proposed exercise lecture 09

3.1 In-class Exercise

Write the observed log-likelihood.
$X_1 \sim \mathcal{N}(\mu_1, \Sigma_{11})$ observed, $X_2 \sim \mathcal{N}(\mu_2, \Sigma_{22})$ missing.

$$\begin{aligned}
\mathcal{L}_{obs}(\theta, X) &= \int \mathcal{L}_{full}(\theta, X) dX^{mis} \\
&= \int \mathbb{P}(X_1, X_2, \theta) dX_2 \\
&= \int \mathbb{P}(X_1, X_2, \theta) dX_2 \\
&= \int \mathbb{P}(X_1)\mathbb{P}(X_2 \mid X_1) dX_2 \\
&= \mathbb{P}(X_1) \int \mathbb{P}(X_2 \mid X_1) dX_2 \\
&= \prod_{i=1}^{n} \mathbb{P}(X_{i1}) \int \mathbb{P}(X_2 \mid X_1) dX_2
\end{aligned}$$

$$\begin{aligned}
\log \mathbb{P}(X_{i1}) &= \frac{n}{2} \log(\sigma_{11}^2) - \frac{1}{2} \sum_{i=1}^{n} \frac{(X_{i1} - \mu_1)^2}{\sigma_{11}^2} \\
&= \frac{n}{2} \log(\sigma_{22} - \frac{\sigma_{21}^2}{\sigma_{11}}) - \frac{1}{2} \sum_{i=1}^{n} \frac{(X_{i2} - (\mu_2 + \frac{\sigma_{21}}{\sigma_{11}}(X_{i1} - \mu_1)))^2}{(\sigma_{22} - \frac{\sigma_{21}^2}{\sigma_{11}})^2}
\end{aligned}$$