Lab 15

Machine Learning 2021-2022 - UMONS
Souhaib Ben Taieb

1

You observe a dataset $\mathscr{D} = \{(\boldsymbol{x}_i, y_i\}_{i=1}^n$ where $\boldsymbol{x}_i \in \mathbb{R}^d$ and $y \in \mathbb{R}$. Assume that your model for the data is

$$y_i \sim \text{Laplace}(\boldsymbol{x}_i^\intercal \boldsymbol{\beta}, 1),$$

where $\boldsymbol{\beta} \in \mathbb{R}^d$ are the parameters of your model, and Laplace(μ, b) is the Laplace distribution with mean μ and scale b. Its probability density function is given by

$$f(y; \mu, b) = \frac{1}{2b} \exp\left(\frac{-|y - \mu|}{b}\right).$$

Write down the formula for the (conditional) log-likelihood as a function of the observed data and the (unknown) parameters $\boldsymbol{\beta}$. Explain your derivations.

2

Consider a three-class classification problem where $X \in [0,1]$ and $Y \in \{0,1,2\}$, with the following data generating process:

$$X \sim U(0,1) \;\text{ and }\; Y|X = x \sim \begin{cases} 0, & \text{with probability } 0.2 \\ 1, & \text{with probability } 0.5x \\ 2, & \text{with probability } 0.8 - 0.5x \end{cases}$$

where $U(a,b)$ is a uniform random variable on the interval $[a,b]$.

(a) What is the expression of the Bayes optimal classifier ?

(b) What is the misclassification error rate of the Bayes optimal classfier ? Your answer should be a scalar.

3

We consider Discriminant Analysis for a one-dimensional two-class classification problem. Let $X \in \mathbb{R}$ be the input variable and $Y \in N, E$, the output. We have the following:

- The prior probabilities are given by $\pi_N = P(Y = N) = \frac{\sqrt{2\pi}}{1+\sqrt{2\pi}}$ and $\pi_E = P(Y = E) = \frac{1}{1+\sqrt{2\pi}}$.
- The distribution of X given $Y = N$ is Gaussian (Normal) with zero mean and variance σ^2, i.e. $X|Y = N \sim \mathcal{N}(0, \sigma^2)$.
- The distribution of X given $Y = E$ is given by:

$$P(X = x|Y = E) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0. \end{cases}$$

Starting from the posterior probabilities $P(Y = N|X = x)$ and $P(Y = E|X = x)$, derive the decision boundary, i.e. an equation in x. Note that only the positive solutions of your equation will be relevant; ignore all $x < 0$.

4

Below is a Principal Component Analysis (PCA) of a dataset *after centering and scaling* each column:

	PC1	PC2	PC3	PC4
X1	-0.5628749	0.2324633	-0.5286078	0.5913599
X2	-0.4214823	-0.6750169	0.5126131	0.3223859
X3	0.5730054	0.2201464	0.3024299	0.7292026
X4	-0.4209386	0.6647170	0.6052584	-0.1209310

	PC1	PC2	PC3	PC4
Standard deviation	1.6165	0.9985	0.50804	0.36313
Cumulative Proportion	0.6533	?	?	?

(a) What is the total variance ?

(b) What proportion of the total variance does the second principal component explain ?

(c) Complete the missing values for the cumulative proportions of total variance explained.

(d) How many principal component directions would we need to explain at least 95% of the variance ?

(e) Let $\phi_1 \in \mathbb{R}^4$ and $\phi_2 \in \mathbb{R}^4$ be the first two loading vectors. What is the value of $\phi_1^\intercal \phi_2$? Briefly explain your answer.