Theoretical Machine Learning
(Theoretical Assignment 2)

Akshat Sharma

May 24, 2024

Problem 1.

(a). We know that X maps from the Ω space to the real space. By definition, the preimage of a set $A \subseteq \mathbb{R}$ under X is

$$X^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \in A\}$$

For A=ϕ, the set of all $\omega \in \Omega \mid X(\omega) \in \phi$ is being considered. For $X(\omega)$ to belong in the empty set, ω cannot take any value(s) in Ω. Therefore, $X^{-1}(A)$ must be an empty set.

(b). $X^{-1}(\mathbb{R}) = \{\omega \in \Omega \mid X(\omega) \in R\}$ For any $\omega \in \Omega, X(\omega) \in \mathbb{R}$ must hold true, from the definition of X. It follows that

$$\omega \in X^{-1}(\mathbb{R}) \; \forall \; \omega \in \Omega$$

Therefore,

$$X^{-1}(\mathbb{R}) \supseteq \Omega$$

Furthermore, for any $\theta \in X^{-1}(\mathbb{R}), X(\theta) \in \mathbb{R}$, implying that $\theta \in \Omega$, since Ω is the Sample Space (X is defined only for the elements of Ω). This yields

$$X^{-1}(\mathbb{R}) \subseteq \Omega$$

Combining the above relations,

$$X^{-1}(\mathbb{R}) = \Omega$$

Problem 2.

We have, the probability distribution function

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-\lambda x} & \text{if } x \geq 0 \end{cases}$$

The support of a random variable is defined as the set of all values ω such that

$$P(x = \omega) > 0$$

i.e., the range of values which x has a nonzero chance of taking. From the definition of the probability distribution it follows that, for $P(x = \omega) > 0$,

$$x \geq 0 \quad and \quad 1 - e^{-\lambda x} > 0 \implies x \geq 0$$

Therefore, the support of X is the set of all non-negative real numbers,

$$\{q \in \mathbb{R} \mid q \geq 0\}$$

Problem 3.

Consider N data points, each of the form (x_i, y_i). We assume a linear model y=mx+c, m and c being the parameters. The L-2 norm loss is then

$$L(m, c) = \frac{1}{N} \sum_{i=1}^{N} (y_i - mx_i - c)^2$$

Consider the parameter matrix: $Z = \begin{pmatrix} m \\ c \end{pmatrix}$. Minimizing the loss function,

$$\frac{dL(m, c)}{dZ} = 0$$

In terms of components,

$$\frac{\partial L}{\partial m} = 0 \quad and \quad \frac{\partial L}{\partial c} = 0$$

Using

$$\frac{\partial L}{\partial m} = \frac{1}{N} \sum_{i=1}^{N} -x_i(y_i - mx_i - c) \quad and \quad \frac{\partial L}{\partial c} = \frac{1}{N} \sum_{i=1}^{N} -(y_i - mx_i - c)$$

we obtain

$$m = \frac{N(\sum xy) - (\sum x)(\sum y)}{N(\sum x^2) - (\sum x)^2} \quad and \quad c = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{N(\sum x^2) - (\sum x)^2}$$

To ascertain that these values of m and c indeed minimize the L2 norm summation, we check whether

$$\frac{d}{dZ^T}(\frac{dL}{dZ}) = \begin{pmatrix} \frac{\partial^2 L}{\partial m^2} & \frac{\partial^2 L}{\partial m \partial c} \\ \frac{\partial^2 L}{\partial m \partial c} & \frac{\partial^2 L}{\partial c^2} \end{pmatrix}$$

is positive definite. This means that for any vector $v \in \mathbb{R}^2$, the following relation must be satisfied:

$$v^T.\frac{d}{dZ^T}(\frac{dL}{dZ}).v > 0 \implies \begin{pmatrix} v_1 & v_2 \end{pmatrix} \begin{pmatrix} \frac{\partial^2 L}{\partial m^2} & \frac{\partial^2 L}{\partial m \partial c} \\ \frac{\partial^2 L}{\partial m \partial c} & \frac{\partial^2 L}{\partial c^2} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} > 0$$

We have the relations,

$$\frac{\partial^2 L}{\partial m^2} = \frac{1}{N} \sum_{i=1}^{N} x_i^2, \quad \frac{\partial^2 L}{\partial c^2} = 1 \quad and \quad \frac{\partial^2 L}{\partial m \, \partial c} = \frac{\partial^2 L}{\partial c \, \partial m} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

Therefore,

$$v^T.\frac{d}{dZ^T}(\frac{dL}{dZ}).v \;=\; \begin{pmatrix} v_1 & v_2 \end{pmatrix} \begin{pmatrix} \frac{\partial^2 L}{\partial m^2} & \frac{\partial^2 L}{\partial m \partial c} \\ \frac{\partial^2 L}{\partial m \partial c} & \frac{\partial^2 L}{\partial c^2} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \;=\; \frac{v_1^2}{N}(\sum x^2) \;+\; \frac{2v_1 v_2}{N}(\sum x) \;+\; v_2^2 > 0;$$

where $(v_1, v_2) \in \mathbb{R}^2$. This is the necessary condition for Z to be a local minimum of the L2 norm loss.