

Machine Learning with Python

MTH786U/P 2023/24

Detailed solutions Coursework 0

Nicola Perra, Queen Mary University of London (QMUL)

Problem 1

1. Compute the gradient ∇L of the function $L : \mathbb{R}_+^2 \rightarrow \mathbb{R}$ defined as

$$L(x, y) = \frac{x}{y} - 1 - \log\left(\frac{x}{y}\right).$$

Here \mathbb{R}_+^2 is the space of all real two-dimensional vectors with positive entries.



Problem 1

1. For the first partial derivative we obtain

$$\frac{\partial L}{\partial x} = \frac{1}{y} - \frac{1}{x} = \frac{x - y}{xy}$$

and for the second one

$$\frac{\partial L}{\partial y} = -\frac{x}{y^2} + \frac{1}{y} = \frac{y - x}{y^2}.$$

Problem 1

1. For the first partial derivative we obtain

$$\frac{\partial L}{\partial x} = \frac{1}{y} - \frac{1}{x} = \frac{x - y}{xy} \longrightarrow d_x \log \frac{x}{y} = \frac{y}{x} \frac{1}{y}$$

and for the second one

$$\frac{\partial L}{\partial y} = -\frac{x}{y^2} + \frac{1}{y} = \frac{y - x}{y^2}.$$

Problem 1

Hence, the entire gradient reads as

$$\nabla L(x, y) = \frac{1}{y} \begin{pmatrix} \frac{x-y}{x} \\ \frac{y-x}{y} \end{pmatrix} .$$

Problem 1

2. Show that L from Question 1 is scalar-invariant, i.e. $L(x, y) = L(cx, cy)$ for any scalar $c > 0$ and all arguments $x > 0, y > 0$.



Problem 1

2. We simply observe

$$\begin{aligned}L(cx, cy) &= \frac{cx}{cy} - 1 - \log\left(\frac{cx}{cy}\right) \\ &= \frac{x}{y} - 1 - \log\left(\frac{x}{y}\right) \\ &= L(x, y),\end{aligned}$$

and, consequently, the function L is scalar-invariant.

Problem 2

1. Compute the expected value \mathbb{E}_x of a (discrete) Poisson-distributed random variable X with probability

$$\rho_x := \exp(-\lambda) \frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \dots, s \quad (1)$$

for a constant $\lambda > 0$. What is the solution for $s \rightarrow \infty$?

Hint: Make use of the identity $\exp(\lambda) = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}$.

Problem 2

The expectation for a discrete Poisson-distributed random variable X reads

$$\mathbb{E}_x[x] = \sum_{x=0}^s x \rho_x = \sum_{x=0}^s x \exp(-\lambda) \frac{\lambda^x}{x!}$$



Problem 2

The expectation for a discrete Poisson-distributed random variable X reads

$$\mathbb{E}_x[x] = \sum_{x=0}^s x \rho_x = \sum_{x=0}^s x \exp(-\lambda) \frac{\lambda^x}{x!}$$

$$\frac{x}{x!} = \frac{x}{x(x-1)(x-2)\dots 1}$$

$$\frac{x}{x!} = \frac{1}{(x-1)!}$$



Problem 2

The expectation for a discrete Poisson-distributed random variable X reads

$$\mathbb{E}_x[x] = \sum_{x=0}^s x \rho_x = \sum_{x=0}^s x \underbrace{\left(\exp(-\lambda) \right)}_{\text{Constant}} \frac{\lambda^x}{x!}$$

$$\frac{x}{x!} = \frac{x}{x(x-1)(x-2)\dots 1}$$
$$\frac{x}{x!} = \frac{1}{(x-1)!}$$



Problem 2

The expectation for a discrete Poisson-distributed random variable X reads

$$\mathbb{E}_x[x] = \sum_{x=0}^s x \rho_x = \sum_{x=0}^s x \underbrace{\left(\exp(-\lambda) \right)}_{\text{Constant}} \frac{\lambda^x}{x!}$$

$$\frac{x}{x!} = \frac{x}{x(x-1)(x-2)\dots 1}$$

$$\frac{x}{x!} = \frac{1}{(x-1)!}$$

$$\lambda^x = \lambda \lambda^{x-1}$$

Problem 2

The expectation for a discrete Poisson-distributed random variable X reads

$$\mathbb{E}_x[x] = \sum_{x=0}^s x \rho_x = \sum_{x=0}^s x \underbrace{\exp(-\lambda)}_{\text{Constant}} \frac{\lambda^x}{x!}$$

$$\frac{x}{x!} = \frac{x}{x(x-1)(x-2)\dots 1}$$

$$\frac{x}{x!} = \frac{1}{(x-1)!}$$

$$= \lambda \exp(-\lambda) \sum_{x=1}^s \frac{\lambda^{x-1}}{(x-1)!}$$

$$\lambda^x = \lambda \lambda^{x-1}$$

Problem 2

$$= \lambda \exp(-\lambda) \sum_{x=1}^s \frac{\lambda^{x-1}}{(x-1)!}$$



Problem 2

$$= \lambda \exp(-\lambda) \sum_{x=1}^s \frac{\lambda^{x-1}}{(x-1)!}$$

Hint: Make use of the identity $\exp(\lambda) = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}$.



Problem 2

$$\begin{aligned} &= \lambda \exp(-\lambda) \sum_{x=1}^s \frac{\lambda^{x-1}}{(x-1)!} \\ &= \lambda \exp(-\lambda) \sum_{x=0}^{s-1} \frac{\lambda^x}{x!}. \end{aligned}$$

Hint: Make use of the identity $\exp(\lambda) = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}$.



Problem 2

$$\begin{aligned} &= \lambda \exp(-\lambda) \sum_{x=1}^s \frac{\lambda^{x-1}}{(x-1)!} \\ &= \lambda \exp(-\lambda) \sum_{x=0}^{s-1} \frac{\lambda^x}{x!}. \end{aligned}$$

Hint: Make use of the identity $\exp(\lambda) = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}$.

Taking the limit $s \rightarrow \infty$ therefore yields

$$\begin{aligned} \mathbb{E}_x[x] &= \lambda \exp(-\lambda) \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \\ &= \lambda \exp(-\lambda) \exp(\lambda) = \lambda. \end{aligned}$$

Problem 2

2. For a uniform (and absolutely continuous) random variable X in $[0, 1]$ compute the expectation of $f(X)$ for

$$f(x) := \begin{cases} -\log(x) & x \in [0, 1/5] \\ 0 & \text{otherwise} \end{cases},$$

Make use of the convention $0 \log(0) = 0$.

Problem 2

$$\mathbb{E}_x[f(x)] = \int_a^b f(x)\rho(x)dx$$



Problem 2

$$\mathbb{E}_x[f(x)] = \int_a^b f(x)\rho(x)dx$$

What's the function $\rho(x)$ here?



Problem 2

For an absolutely continuous uniform random variable on the interval $[a, b]$ the probability density function reads

$$\rho(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & \text{otherwise} \end{cases} .$$



Problem 2

For an absolutely continuous uniform random variable on the interval $[a, b]$ the probability density function reads

$$\rho(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & \text{otherwise} \end{cases} .$$

Hence, for $[a, b] = [0, 1]$ we simply have $\rho(x) = 1$ for $x \in [0, 1]$, and we compute

Problem 2

For an absolutely continuous uniform random variable on the interval $[a, b]$ the probability density function reads

$$\rho(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & \text{otherwise} \end{cases} .$$

Hence, for $[a, b] = [0, 1]$ we simply have $\rho(x) = 1$ for $x \in [0, 1]$, and we compute

$$\begin{aligned} \mathbb{E}_x[f(x)] &= \int_0^1 f(x) dx = - \int_0^{\frac{1}{5}} \log(x) dx = - [x \log x - x]_0^{\frac{1}{5}} \\ &= \frac{1}{5} - \frac{1}{5}(\log(1) - \log(5)) = \frac{1}{5} (1 + \log(5)) \approx 0.5218875825 . \end{aligned}$$

Problem3

1. Let X be a random variable with expectation μ and variance σ^2 . Show that the variance of $aX + b$, where $a, b \in \mathbb{R}$, is

$$\text{Var}_x[ax + b] = a^2\sigma^2.$$



Problem3

1. With the definition of the variance we compute

$$\text{Var}_x[ax + b] = \mathbb{E}_x [(ax + b - \mathbb{E}_x[ax + b])^2]$$



Problem3

1. With the definition of the variance we compute

$$\begin{aligned}\text{Var}_x[ax + b] &= \mathbb{E}_x [(ax + b - \mathbb{E}_x[ax + b])^2] \\ &= \mathbb{E}_x [(ax + b - \mathbb{E}_x[ax] - \mathbb{E}_x[b])^2]\end{aligned}$$



Problem3

1. With the definition of the variance we compute

$$\begin{aligned}\text{Var}_x[ax + b] &= \mathbb{E}_x [(ax + b - \mathbb{E}_x[ax + b])^2] \\ &= \mathbb{E}_x [(ax + b - \mathbb{E}_x[ax] - \mathbb{E}_x[b])^2] \\ &= \mathbb{E}_x [(ax + b - a\mathbb{E}_x[x] - b\mathbb{E}_x[1])^2]\end{aligned}$$



Problem3

1. With the definition of the variance we compute

$$\begin{aligned}\text{Var}_x[ax + b] &= \mathbb{E}_x [(ax + b - \mathbb{E}_x[ax + b])^2] \\ &= \mathbb{E}_x [(ax + b - \mathbb{E}_x[ax] - \mathbb{E}_x[b])^2] \\ &= \mathbb{E}_x [(ax + b - a\mathbb{E}_x[x] - b\mathbb{E}_x[1])^2] \\ &= \mathbb{E}_x [(ax + b - a\mathbb{E}_x[x] - b)^2]\end{aligned}$$



Problem3

1. With the definition of the variance we compute

$$\begin{aligned}\text{Var}_x[ax + b] &= \mathbb{E}_x [(ax + b - \mathbb{E}_x[ax + b])^2] \\ &= \mathbb{E}_x [(ax + b - \mathbb{E}_x[ax] - \mathbb{E}_x[b])^2] \\ &= \mathbb{E}_x [(ax + b - a\mathbb{E}_x[x] - b\mathbb{E}_x[1])^2] \\ &= \mathbb{E}_x [(ax + b - a\mathbb{E}_x[x] - b)^2] \\ &= \mathbb{E}_x [(ax - a\mathbb{E}_x[x])^2]\end{aligned}$$



Problem3

1. With the definition of the variance we compute

$$\begin{aligned}\text{Var}_x[ax + b] &= \mathbb{E}_x [(ax + b - \mathbb{E}_x[ax + b])^2] \\ &= \mathbb{E}_x [(ax + b - \mathbb{E}_x[ax] - \mathbb{E}_x[b])^2] \\ &= \mathbb{E}_x [(ax + b - a\mathbb{E}_x[x] - b\mathbb{E}_x[1])^2] \\ &= \mathbb{E}_x [(ax + b - a\mathbb{E}_x[x] - b)^2] \\ &= \mathbb{E}_x [(ax - a\mathbb{E}_x[x])^2] \\ &= \mathbb{E}_x [a^2 (x - \mathbb{E}_x[x])^2]\end{aligned}$$

Problem3

1. With the definition of the variance we compute

$$\begin{aligned}\text{Var}_x[ax + b] &= \mathbb{E}_x [(ax + b - \mathbb{E}_x[ax + b])^2] \\ &= \mathbb{E}_x [(ax + b - \mathbb{E}_x[ax] - \mathbb{E}_x[b])^2] \\ &= \mathbb{E}_x [(ax + b - a\mathbb{E}_x[x] - b\mathbb{E}_x[1])^2] \\ &= \mathbb{E}_x [(ax + b - a\mathbb{E}_x[x] - b)^2] \\ &= \mathbb{E}_x [(ax - a\mathbb{E}_x[x])^2] \\ &= \mathbb{E}_x [a^2 (x - \mathbb{E}_x[x])^2] \\ &= a^2 \mathbb{E}_x [(x - \mathbb{E}_x[x])^2]\end{aligned}$$

Problem3

1. With the definition of the variance we compute

$$\begin{aligned}\text{Var}_x[ax + b] &= \mathbb{E}_x [(ax + b - \mathbb{E}_x[ax + b])^2] \\ &= \mathbb{E}_x [(ax + b - \mathbb{E}_x[ax] - \mathbb{E}_x[b])^2] \\ &= \mathbb{E}_x [(ax + b - a\mathbb{E}_x[x] - b\mathbb{E}_x[1])^2] \\ &= \mathbb{E}_x [(ax + b - a\mathbb{E}_x[x] - b)^2] \\ &= \mathbb{E}_x [(ax - a\mathbb{E}_x[x])^2] \\ &= \mathbb{E}_x [a^2 (x - \mathbb{E}_x[x])^2] \\ &= a^2 \mathbb{E}_x [(x - \mathbb{E}_x[x])^2] \\ &= a^2 \text{Var}_x[x]\end{aligned}$$