IFT 6390 Fundamentals of Machine Learning
Ioannis Mitliagkas

Homework 0

1 Theoretical Part [5 points]

1. [1 points] Let X be a random variable representing the outcome of a single roll of a 6-sided dice. Show the steps for the calculation of i) the expectation of X and ii) the variance of X.

2. [1 points] Let $u, v \in \mathbb{R}^d$ be two vectors and let $A \in \mathbb{R}^{n\times d}$ be a matrix. Give the formulas for the euclidean norm of u, for the euclidean inner product (aka dot product) between u and v, and for the matrix-vector product Au.

3. [1 points] Consider the two algorithms below. What do they compute and which algorithm is faster?
Observez les deux algorithms ci-dessous. Que calculent-ils et lequel est le plus rapide ?

ALGO1(n)
```
result = 0
for i = 1...n
  result = result + i
return result
```

ALGO2(n)
```
return (n+1)*n/2
```

4. [1 points] Give the step-by-step derivation of the following derivatives:

$$\text{i)} \quad \frac{df}{dx} =?, \quad \text{where } f(x,\beta) = x^2 \exp(-\beta x)$$

$$\text{ii)} \quad \frac{df}{d\beta} =?, \quad \text{where } f(x,\beta) = x \exp(-\beta x)$$

$$\text{iii)} \quad \frac{df}{dx} =?, \quad \text{where } f(x) = \sin(\exp(x^2))$$

5. [1 points] Let $X \sim N(\mu, 1)$, that is the random variable X is distributed according to a Gaussian with mean μ and standard deviation 1. Show how you can calculate the second moment of X, given by $\mathbb{E}[X^2]$.

Solutions

1. (a) Expectation of a discrete random variable X can be defined as:

$$\mathbf{E}[X] = \sum_i x_i * p(x_i) \tag{1}$$

In the given problem, X can take values from the discrete set of {1,2,3,4,5,6} with probability $p(x_i) = 1/6$ for each event.

Thus $\mathbf{E}[X] = [1*1/6+2*1/6+...6*1/6]$, which gives $\mathbf{E}[X] = 3.5$ as the answer.

(b) Variance of a discrete random variable X can be defined as:

$$\mathbf{Var}[X] = \sum_i (x_i - \mu)^2 * p(x_i) \tag{2}$$

where μ represents the mean or $\mathbf{E}[X] = 3.5$. Substituting the values in above equation gives:

$\mathbf{Var}[X] = [(1-3.5)^2 * 1/6 + (2-3.5)^2 * 1/6 + ..(6-3.5)^2 * 1/6]$

$\mathbf{Var}[X] = 17.5/6$

2. Let vector u be defined as $= \begin{bmatrix} u_1 & u_2 & .. & u_d \end{bmatrix}$ and $v = \begin{bmatrix} v_1 & v_2 & .. & v_d \end{bmatrix}$

(a) Euclidean norm of u is defined as: $||u||_2 = \sqrt{\sum_{i=1}^{d} u_i^2}$

(b) Dot product of u and v is defined as:

$$u \cdot v = \sum_{i=1}^{d} u_i * v_i$$

(c) The matrix vector product $A \cdot u = \begin{bmatrix} A_{11} & A_{12} & .. & A1d \\ A_{21} & A_{22} & .. & A_{2d} \\ .. & .. & .. & .. \\ A_{n1} & A_{n2} & .. & A_{nd} \end{bmatrix}_{nxd} \begin{bmatrix} u_1 \\ u_2 \\ .. \\ u_d \end{bmatrix}_{dx1}$

$$A \cdot u = \begin{bmatrix} A_{11}u_1 + A_{12}u_2 + .. + A1du_d \\ A_{21}u_1 + A_{22}u_2 + .. + A_{2d}u_d \\ .. \\ A_{n1}u_1 + A_{n2}u_2 + .. + A_{nd}u_d \end{bmatrix}_{nx1}$$

3. Both algorithms compute the sum of integers between 1 to n. The time complexity of **ALGO1** $= \mathcal{O}(n)$ and **ALGO2** $= \mathcal{O}(1)$. Hence **ALGO2** is faster.

4. (a) For the function $f(x, \beta) = x^2 \exp(-\beta x)$, let $u(x) = x^2$ and $v(x) = \exp(-\beta x)$.

$$\frac{df}{dx} = \frac{du}{dx} * v + u * \frac{dv}{dx} \text{ using the product rule of differentiation}$$

$$\frac{df}{dx} = 2x * \exp(-\beta x) + x^2 * \exp(-\beta x) * (-\frac{d\beta}{dx}x - \beta)$$

$$\frac{df}{dx} = \exp(-\beta x)(2x - \beta x^2 - \frac{d\beta}{dx}x)$$

(b) $f(x, \beta) = x \exp(-\beta x)$

$$\frac{df}{d\beta} = \frac{dx}{d\beta} \exp(-\beta x) + x \exp(-\beta x)(-x - \beta \frac{dx}{d\beta})$$

(c) $f(x) = \sin(\exp(x^2))$

$$\frac{df}{dx} = \cos(\exp(x^2)) * \exp(x^2) * 2x$$

5. The $\mathbf{Var}[X]$ of normally distributed random variable X is defined as:

$\mathbf{Var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2]$, where mean $\mu = \mathbf{E}[X]$

$\mathbf{Var}[X] = \mathbf{E}[(X^2 + \mathbf{E}[X]^2 - 2X\mathbf{E}[X])]$

using the linearity property of Expectation, we can expand the above term as: $\mathbf{Var}[X] = \mathbf{E}[X^2] + \mathbf{E}[X]^2 - 2\mathbf{E}[X]\mathbf{E}[X]$

$\mathbf{Var}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2$

thus, $\mathbf{E}[X^2] = \mathbf{Var}[X] + \mathbf{E}[X]^2$

since in given distribution, $\mathbf{Var}[X] = 1$ and $\mathbf{E}[X] = \mu$

therefore, $\mathbf{E}[X^2] = 1 + \mu^2$