IFT-3395/6390 Fundamentals of Machine Learning
Professors: Guillaume Rabusseau / Ioannis Mitliagkas

Homework 0

by Tony Brière

1. Undergraduates 1 pts Graduates 1 pts
   Question. 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.

   Answer. Let $p(x)$ be the probability mass function of X.

$$p(x) = \begin{cases} 1/6, & x \in \{1, 2, 3, 4, 5, 6\} \\ 0, & otherwise \end{cases}$$

   (i)

$$\begin{aligned} E(X) &= \sum_{x=1}^{6} xp(x) \\ &= \sum_{x=1}^{6} 1/6x \\ &= 1/6 \sum_{x=1}^{6} x \\ &= 1/6 \times 6 \times (6+1)/2 \\ &= 3.5 \end{aligned}$$

(ii)

$$
\begin{aligned}
E(X^2) &= \sum_{x=1}^{6} x^2 p(x) \\
&= \sum_{x=1}^{6} 1/6 x^2 \\
&= 1/6 \sum_{x=1}^{6} x^2 \\
&= 1/6 \times \frac{6 \times (6+1) \times (2 \times 6 + 1)}{6} \\
&= 15.1\bar{6}
\end{aligned}
$$

$$
\begin{aligned}
Var(X) &= E(X^2) - E^2(X) \\
&= 15.1\bar{6} - 3.5^2 \\
&= 2.91\bar{6}
\end{aligned}
$$

2. Undergraduates 1 pts Graduates 1 pts
   Question. 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.

   Answer.

   (a) Euclidean norm:

   $$\|u\| = \sqrt{\sum_{i=1}^{d} u_i^2}$$

   (b) Euclidean inner product:

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

   (c) Matrix-vector product:

   $$(Au)_i = \sum_{j=1}^{d} A_{i,j} u_j$$

3. Undergraduates 1 pts Graduates 1 pts
   Question. Consider the two algorithms below. What do they compute and which algorithm is faster?

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

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

   Answer. Both algorithms are computing the sum of the first n integers. The first algorithm has give or take $2n+3$ instructions. It is $O(n)$ whereas the second algorithm has 4 instructions (including return) regardless of n. The second algorithm is $O(1)$ and it is faster than the first $\forall n > 0$. When $n = 0$, the first algorithm is one instruction faster.

4. Undergraduates 1 pts Graduates 1 pts
   Question. Give the step-by-step derivation of the following derivatives:

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

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

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

   Answer.

   (i)

$$\begin{aligned}
\frac{d(x^2 \exp(-\beta x))}{dx} &= x^2 \frac{d(\exp(-\beta x))}{dx} + \exp(-\beta x)\frac{d(x^2)}{dx} \\
&= x^2 \exp(-\beta x)\frac{d(-\beta x)}{dx} + 2x \exp(-\beta x) \\
&= -\beta x^2 \exp(-\beta x) + 2x \exp(-\beta x) \\
&= x \exp(-\beta x)[2 - \beta x]
\end{aligned}$$

(ii)

$$\begin{aligned}\frac{d(x\exp(-\beta x))}{d\beta} &= x\frac{d(\exp(-\beta x))}{d\beta}\\ &= x\exp(-\beta x)\frac{d(-\beta x)}{d\beta}\\ &= x\exp(-\beta x)(-x)\\ &= -x^2\exp(-\beta x)\end{aligned}$$

(iii)

$$\begin{aligned}\frac{d(\sin(\exp(x^2)))}{dx} &= \cos(\exp(x^2))\frac{d(\exp(x^2))}{dx}\\ &= \cos(\exp(x^2))\exp(x^2)\frac{d(x^2)}{dx}\\ &= \cos(\exp(x^2))\exp(x^2)(2x)\\ &= 2x\exp(x^2)\cos(\exp(x^2))\end{aligned}$$

5. Undergraduates 1 pts Graduates 1 pts
Question. 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 $\mathbb{E}[X^2]$.

Answer.

$$\begin{aligned}Var(X) &= E[(X - E[X])^2]\\ &= E[X^2 - 2XE[X] + E^2[X]]\\ &= E[X^2] - 2E[XE[X]] + E^2[X]\\ &= E[X^2] - 2E[X]E[X] + E^2[X]\\ &= E[X^2] - 2E^2[X] + E^2[X]\\ &= E[X^2] - E^2[X]\end{aligned}$$

$$\begin{aligned}\Rightarrow E[X^2] &= Var(X) + E^2[X]\\ &= 1^2 + \mu^2\\ &= \mu^2 + 1\end{aligned}$$