Introduction to Deep Learning for Speech and Language Processing
Exercise Sheet 3: Math for Machine Learning

Thang Vu

29th October 2025

Probabilities

Exercise 1.
What is the probability of throwing a fair dice and getting a number greater than 3?
Solution 1.

- Sample space Ω: 6 possible outcomes
- Event space A includes 4, 5, 6, i.e. 3 possible outcomes
- $P(A) = 3/6$

Exercise 2.
What is the probability of throwing two fair dices and getting a sum that is greater than 8?
Solution 2.

- Sample space Ω: 36 possible outcomes
- Event space A includes(3, 6), (4, 5), (4, 6), (5, 4), (5, 5), (5, 6), (6, 3), (6, 4), (6, 5), (6,6), i.e. 10 possible outcomes
- $P(A) = 10/36$

Exercise 3.
Given two random variables X and Y and the following table that summarizes how many times the two events occur:

Y/X	x_1	x_2	x_3	x_4
y_1	4	3	5	10
y_2	1	8	3	2

Your tasks are:

- Compute the joint probability $P(X = x_2, Y = y_1)$ and $P(Y = y_2, X = x_1)$
- Compute the marginal probability $P(X = x_2)$
- Compute the marginal probability $P(Y = y_1)$
- Compute the conditional probability $P(X = x_3 \mid Y = y_2$
- Compute the conditional probability $P(Y = y_2 \mid X = x_3$

Solution 3.

- $N = 36$
- $P(X = x_2, Y = y_1) = 3/36$ and $P(Y = y_2, X = x_1) = 1/36$
- $P(X = x_2) = 11/36$
- $P(Y = y_1) = 22/36$
- $P(X = x_3 \mid Y = y_2 = 3/14$
- $P(Y = y_2 \mid X = x_3 = 3/8$

Exercise 4.

We happen to have a dice where its faces do not show up equally. Let X be the random variable representing the value of the dice after rolling and we know that:

- P(X = 1) = 0.1
- P(X = 2) = 0.1
- P(X = 3) = 0.1
- P(X = 4) = 0.2
- P(X = 5) = 0.2
- P(X = 6) = 0.3

You tasks are:

- Compute the expected value $E[X]$.
- Calculate the variance $Var[X]$.

Solution 4.

- $E[X] = 1 \cdot 0.1 + 2 \cdot 0.1 + 3 \cdot 0.1 + 4 \cdot 0.2 + 5 \cdot 0.2 + 6 \cdot 0.3 = 4.2$
- $Var[X] = (1-4.2)^2 \cdot 0.1 + (2-4.2)^2 \cdot 0.1 + (3-4.2)^2 \cdot 0.1 + (4-4.2)^2 \cdot 0.2 + (5-4.2)^2 \cdot 0.2 + (6-4.2)^2 \cdot 0.3 = 2.76$

Exercise 5.

Let's consider two events that are dependent from each other. The events will either take place ('1') or be canceled ('0'). Let X_1 be the random variable representing the outcome of the first event and X_2 be the random variable representing the outcome of the second event. The following table presents the joint probability of these two events:

x_1	x_2	$P(X_1 = x_1, X_2 = x_2)$
0	0	0.2
0	1	0.1
1	0	0.1
1	1	0.6

You tasks are:

- Compute the expected values $E[X_1]$ and $E[X_2]$.
- Calculate the covariance $Cov(X_1, X_2)$ between the two events.

Solution 5.

- $E[X_1] = 0 \cdot 0.2 + 0 \cdot 0.1 + 1 \cdot 0.1 + 1 \cdot 0.6 = 0.7$
- $E[X_1] = 0 \cdot 0.2 + 1 \cdot 0.1 + 0 \cdot 0.1 + 1 \cdot 0.6 = 0.7$
- $Cov(X_1, X_2) = (0-0.7) \cdot (0-0.7) \cdot 0.2 + (0-0.7) \cdot (1-0.7) \cdot 0.1 + (1-0.7) \cdot (0-0.7) \cdot 0.1 + (1-0.7) \cdot (1-0.7) \cdot 0.6 = 0.11$

Optimization

Exercise 6.

(1) Given $f(x) = (x+2)^2 + 3$ with $x \in \mathbb{R}$ perform one step of gradient descent starting from $x^0 = 1$ with a learning $\eta = 0.1$.

(2) Given $f(x) = x_1^3 + 2x_2^2 - 1$ with $x \in \mathbb{R}^2$ perform one step of gradient descent starting from $x^0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ with a learning $\eta = 0.1$.

Solution 6.

- $f'(x) = 2x + 4$; $f'(x = 1) = 6$; $x^1 = x^0 - 0.1 \cdot 6 = 0.4$
- $\dfrac{\partial f(x)}{\partial x} = \begin{bmatrix} 3x_1^2 \\ 4x_2 \end{bmatrix}$; $\dfrac{\partial f(\begin{bmatrix} 1 \\ 0 \end{bmatrix})}{\partial x} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$; $x^1 = x^0 - 0.1 \begin{bmatrix} 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} - 0.1 \begin{bmatrix} 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.7 \\ 0 \end{bmatrix}$