

# Solution for ML F16 assignment 1: Math and ML

Jonathan Hansen fdg890

November 28, 2016

## 1 Vectors and Matrices

To help solve the questions of this task I have used the book "Linear Algebra for Engineers and Scientists" [1]. We are provided with the following vectors:

$$a = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$$

and the matrix:

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

### Question 1.1

The inner product of  $a$  and  $b$  is

$$a \cdot b = 1 \cdot 3 + 2 \cdot 2 + 2 \cdot 1 = 3 + 4 + 2 = 9$$

### Question 1.2

The Euclidean norm of vector  $a$  is

$$\|a\| = \sqrt{1^2 + 2^2 + 2^2} = \sqrt{9} = 3$$

### Question 1.3

The outer product of  $a$  and  $b$  is

$$ab^T = \begin{bmatrix} 1 \cdot 3 & 1 \cdot 2 & 1 \cdot 1 \\ 2 \cdot 3 & 2 \cdot 2 & 2 \cdot 1 \\ 2 \cdot 3 & 2 \cdot 2 & 2 \cdot 1 \end{bmatrix} = \begin{bmatrix} 3 & 2 & 1 \\ 6 & 4 & 2 \\ 6 & 4 & 2 \end{bmatrix}$$

### Question 1.4

According to [1] **Definition 4.1** dot products are commutative why  $b^t a = a^t b$ .

### Question 1.5

The inverse matrix of  $M$  can be found by reducing the partitioned matrix made up of  $M$  and the 3x3 identity matrix:

$$\mathbf{M}^{-1} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 0 & 0 & 1 \end{bmatrix} \sim \begin{matrix} \\ r_2 \cdot 1/4 \rightarrow r_2 \\ r_3 \cdot 1/2 \rightarrow r_3 \end{matrix} \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1/4 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1/2 \end{bmatrix}$$

This gives us the inverse matrix:

$$\mathbf{M}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/4 & 0 \\ 0 & 0 & 1/2 \end{bmatrix}$$

### Question 1.6

The matrix-vector product  $Ma$  is equal to

$$\mathbf{M}\mathbf{a} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{bmatrix} \cdot \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} = \begin{bmatrix} 1 \cdot 1 & 0 \cdot 2 & 0 \cdot 2 \\ 1 \cdot 0 & 2 \cdot 4 & 2 \cdot 0 \\ 1 \cdot 0 & 2 \cdot 0 & 2 \cdot 2 \end{bmatrix} = \begin{pmatrix} 1 \\ 8 \\ 4 \end{pmatrix}$$

### Question 1.7

The transpose of  $\mathbf{A}$  is formed by writing the columns of  $\mathbf{A}$  as the rows of  $\mathbf{A}^T$ :

$$\mathbf{A}^T = \begin{bmatrix} 3 & 6 & 6 \\ 2 & 4 & 4 \\ 1 & 2 & 2 \end{bmatrix}$$

It is clear to see, that  $\mathbf{A}$  is not symmetric.

### Question 1.8

A reduction of  $\mathbf{A}$  to echelon form would leave just 1 non-zero row, since all 3 rows are linear combinations of the row (1 2 2). For the same reason, the rank of  $\mathbf{A}$  is 1.

### Question 1.9

According to **Theorem 3.5** of [1] a square matrix is invertible if and only if its columns are linearly independent. Combined with **Theorem 3.4** that states that the rank of a  $n \times k$  matrix is equal to the number of columns if they are independent we get the answer to our question: In for a square matrix to be invertible the number of columns should be equal to the rank and the columns should be invertible. For the same reason  $\mathbf{A}$  is not invertible.

## 2 Derivatives

I have used the book "Kalkulus" [2] to help answer the questions in this task.

**Question 2.1**

Begin by rewriting  $f(x)$  to make it easier to work with and use that

$$\exp(-x) = e^{-x}$$

So we are looking for the derivative of

$$f(x) = \frac{1}{e^x + 1}$$

And so we get:

$$\frac{df(x)}{dx} = \frac{e^{-x}}{(e^{-x} + 1)^2}$$

**Question 2.2**

The partial derivative of  $f(w, x)$  with respect to  $w$  is

$$\frac{\partial}{\partial w} = 4x(xw + 5)$$

### 3 Probability Theory: Sample Space

I have used the book "Introduction to Probability" [3] to help answer the questions in this task.

**Question 3.1**

The sample space is the set of all possible outcomes of an experiment. In this case where we draw two balls of an urn then the outcome is a set of two balls. Notice that in the answer below the order of the balls of a pair is insignificant meaning that for instance the pair (red, orange) is equal to the pair (orange, red). So the sample space is:

$$S = \{(red, red), (red, orange), (red, blue), (orange, orange), (orange, blue)\}$$

### Question 3.2

To calculate the probabilities of the points in the sample space we can think of an outcome as made up of two events where one ball is picked in each event. Notice that due to the chosen insignificance of the order of balls in a pair this means that for instance the pair (red, orange) can happen both by first picking a red ball or by first picking an orange. This is important in order for the sum of the probabilities of the points to sum to 1.

The chance of picking a red ball as the first is  $\frac{5}{9}$  since there is a total of 9 balls and 5 of them are red. To make a pair we have to pick another ball, but since we already picked one the total number of balls is now 8. There are only 4 red balls left, but the number of orange and blue balls is the same. The chance of picking another red is then  $\frac{4}{8}$  and the total probability of the (red, red) outcome is the product of the probability of these two single events. Below  $\times 2$  denotes these mirror outcomes. Using the explained logic we get the following probabilities:

$$P(\text{red}, \text{red}) = \frac{5}{9} \cdot \frac{4}{8} = \frac{20}{72}$$

$$P(\text{red}, \text{orange}) = \frac{5}{9} \cdot \frac{3}{8} = \frac{15}{72} \quad \times 2$$

$$P(\text{red}, \text{blue}) = \frac{5}{9} \cdot \frac{1}{8} = \frac{5}{72} \quad \times 2$$

$$P(\text{orange}, \text{orange}) = \frac{3}{9} \cdot \frac{2}{8} = \frac{6}{72}$$

$$P(\text{orange}, \text{blue}) = \frac{3}{9} \cdot \frac{1}{8} = \frac{3}{72} \quad \times 2$$

### Question 3.3

There are three orange balls in the urn. Since we only need to pick two balls it is possible to get the outcome, that they are both orange. There are other balls in the urn as well so we can also have zero orange balls and an outcome with one orange and some other ball. Hence the possible values of  $\mathbf{X}$  are:

$$\mathbf{X} = [0, 1, 2]$$

### Question 3.4

From question 3.2 we know the possible outcomes of the experiment (the points) and their probabilities. There is two possible outcomes without any orange balls: A pair of reds and a pair of a red and a blue. Thus we get the following probability:

$$\mathbb{P}\{\mathbf{X} = 0\} = \frac{20}{72} + \frac{5}{72} + \frac{5}{72} = \frac{30}{72}$$

### Question 3.5

According to [3] **Definition 4.1.1** the expected value of a discrete variable is the weighted average of the values that the variable can take on. So we get:

$$\mathbb{E}[\mathbf{X}] = 0 \cdot \frac{30}{72} + 1 \cdot \frac{15 + 15 + 3 + 3}{72} + 2 \cdot \frac{6}{72} = \frac{48}{72} = \frac{2}{3}$$

## 4 Probability Theory: Properties of Expectation

The assignment took me longer to finish than I planned for, so unfortunately I did not do this task.

## 5 Probability Theory: Complements of Events

I have used the book "Introduction to Probability" [3] for this task. Especially section 1.6 was useful here.

### Question 5.1

Begin by realizing that the complementary event  $\bar{A}$  will actually often be what we intuitively would understand as a set of events. In the case of a coin flip if we do not get for instance heads, we will get just tails, but as soon as there is more than two possible outcomes the chance of one event not happening will be equal to the sum of the probabilities of all the other events in the sample space happening. The first probability axiom states that the probability of an entire sample space is equal to one,  $P(S) = 1$ . The second axiom states that if we have disjoint events of the sample space then their combined probability is the sum of their individual probabilities. The complementary event of  $\mathbf{A}$  is defined exactly as "**not A**" why the two are mutually exclusive, which in terms means disjoint. Combining our intuition with the axioms it is then clear, that if the probability of the entire sample space has to be 1 and we know the combined probability of all the events **not A**, then the chance of  $\mathbf{A}$  must be exactly what is left,  $\mathbb{P}\{A\} = 1 - \mathbb{P}\{\bar{A}\}$ .

### Question 5.2

Flipping a coin 10 times can result in only one outcome where there is not at least one tail: The outcome of all heads. The chance of getting this is

$$P(\text{All heads}) = \frac{1}{2^{10}} = \frac{1}{1024}$$

From the question above we know then that getting all other outcomes, in this case all with at least one tails, is equal to

$$P(\text{At least one tail}) = 1 - \frac{1}{1024} = \frac{1023}{1024}$$

To find the probability of observing at least two tails we can use the binomial PMF combined with the procedure from above. We want to find out what the probability is of getting just zero or one tail, since all other outcomes will have at least two tails:

$$P(\text{At least two tails}) = 1 - P(\text{Zero tails}) - P(\text{One tail}) = 1 - (1 - 0.5)^{10} - 10 \cdot 0.5 \cdot (1 - 0.5)^9 = 0.99$$

## 6 Probability Theory: Coin Flips

I have used the book "Introduction to Probability" [3] for this task.

### Question 6.1

For the number of heads and tails to be equal there must be 5 of each. Since the coin is fair the computation is the same whether we do it for heads or tails. Again we can use the binomial PMF:

$$P(\text{Equal heads and tails}) = \binom{10}{5} 0.5^5 (1 - 0.5)^5 = 0.25$$

### Question 6.2

For there to be more heads than tails any outcome with at least 6 heads is good. As with Question 5.2.2 it is actually easier to find the combined chance that we get 0, 1, 2, 3, and 4 tails, since this time there is not just one exact number of heads that satisfies the question:

$$\begin{aligned} P(\text{Heads} > \text{Tails}) &= 1 - P(0 \text{ tails}) - P(1 \text{ tail}) - P(2 \text{ tails}) - P(3 \text{ tails}) - P(4 \text{ tails}) \\ &= 0.38 \end{aligned}$$

### Question 6.3

I am not sure how to solve this. I have an intuition that the probability we are looking for is surprisingly large. My intuition is that since we always flip 10 coins any  $i$ -th flip will also have an  $(11 - i)$ -th flip and there is a 50 % chance that any two flips are the same. So my guess is, that in total there is also 50 % chance to get the event asked for. However my intuition on probability is scandalous, so I am probably (aha!) wrong.

## References

- [1] K. Hardy, *Linear Algebra for Engineers and Scientists*. Pearson Prentice Hall, 2005, vol. 1.
- [2] T. Lindstrom, *Kalkulus*. universitetsforlaget, 2006, vol. 1.
- [3] J. K. Blitzstein and J. Hwang, *Introduction to Probability*. CRC Press, 2015, vol. 1.