Machine Learning 2016 - Assignment 1

Christoffer Thrysøe - dfv107

February 20, 2017

1. Vectors and Matrices

For this sub problem the following vectors and matrix were given :

$$a = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \; b = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} \; and \; \mathbf{M} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Question 1

Compute $a^T b$:

$$\begin{bmatrix} 1 & 2 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} = (1 \times 3) + (2 \times 2) + (2 \times 1) = 9$$

thus, $a^T b = 9$

Question 2

Compute the length of vector **a**:

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

thus, $\|a\| = 3$

Question 3

$$\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \begin{bmatrix} 3 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 3 \times 1 & 2 \times 1 & 1 \times 1 \\ 3 \times 2 & 2 \times 2 & 1 \times 2 \\ 3 \times 2 & 2 \times 2 & 1 \times 2 \end{bmatrix} = \begin{bmatrix} 3 & 2 & 1 \\ 6 & 4 & 2 \\ 6 & 4 & 2 \end{bmatrix}$$

Therefore $a^T b \neq ab^T$ so the answer is "No".

Question 4

$$\begin{bmatrix} 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = (3 \times 3) + (2 \times 2) + (1 \times 2) = 9$$

Therefore $a^T b = b^T a$, and the answer is "Yes".

Question 5

To calculate the inverse of matrix M, the matrix is brought to row echelon form, the same operations are applied to an Identity matrix I. The result of applying the operations to I will be the inverse of matrix M, if reduction is possible.

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 0 & 0 & 1 \end{array}\right]$$

Row 2 is multiplied by $\frac{1}{4}$, row 3 is multiplied by $\frac{1}{2}$. The resulting matrix after the row operations are:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 1 & 0 & 0 & \frac{1}{2} \end{array}\right]$$

Therefore:

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

Question 6

Compute **Ma**:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} (1 \times 1) + (1 \times 0) + (1 \times 0) \\ (2 \times 0) + (2 \times 4) + (2 \times 0) \\ 2 \times 0) + (2 \times 0) + (2 \times 2) \end{bmatrix} = \begin{bmatrix} 1 \\ 8 \\ 4 \end{bmatrix}$$

Therefore:

$$\mathbf{Ma} = \begin{bmatrix} 1 \\ 8 \\ 4 \end{bmatrix}$$

Question 7

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

A is not symmetric as $\mathbf{A} \neq \mathbf{A^T}$.

Question 8

The matrix **A** can be reduced to row echelon form, by subtracting two times row 1 from row 2 and one time row 2 from row 3.

$$\begin{bmatrix} 3 & 2 & 1 \\ 6 & 4 & 2 \\ 6 & 4 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} 3 & 2 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The number of linearly independent rows in the reduced matrix is 1, therefore $\mathbf{Rank}(\mathbf{A}) = 1$

Question 9

In order for a square matrix to be invertible, it must have full rank, that is for a matrix of size $m \times m$ the rank must be m.

2. Derivatives

Question 1

$$
\begin{aligned}
f(x) &= \frac{1}{e^{-x}+1} \\
\frac{\partial}{\partial x} f(x) &= \frac{\partial}{\partial x}\left[\frac{1}{e^{-x}+1}\right] \\
&= -\frac{\frac{\partial}{\partial x}\left[e^{-x}+1\right]}{(e^{-x}+1)^2} && \textit{(Reciprocal rule)} \\
&= -\frac{\frac{\partial}{\partial x}\left[e^{-x}\right]+\frac{\partial}{\partial x}\left[1\right]}{(e^{-x}+1)^2} \\
&= -\frac{e^{-x}\frac{\partial}{\partial x}\left[-x\right]+0}{(e^{-x}+1)^2} && \textit{(Exponential function rule)} \\
&= -\frac{e^{-x}-\frac{\partial}{\partial x}\left[x\right]}{(e^{-x}+1)^2} \\
&= \frac{e^{-x}}{(e^{-x}+1)^2}
\end{aligned}
$$

Question 2

$$
\begin{aligned}
f(w,x) &= 2(wx+5)^2 \\
\frac{\partial}{\partial w} f(w,x) &= \frac{\partial}{\partial w}\left[2(wx+5)^2\right] \\
&= 2\frac{\partial}{\partial w}\left[(wx+5)^2\right] \\
&= 2 \times 2(wx+5)\frac{\partial}{\partial w}\left[wx+5\right] && \textit{(Power rule)} \\
&= 4(x\frac{\partial}{\partial w}\left[w\right] + \frac{\partial}{\partial w}\left[5\right])(wx+5) \\
&= 4(1x+0)(wx+5) \\
&= 4x(wx+5)
\end{aligned}
$$

3. Probability Theory: Sample Space

Question 1

The sample space Ω is the set of possible outcomes.
There are three different colours, 9 balls. 2 balls are picked from the urn uniformly at random. In the urn there is 5 red balls, 3 orange balls and 1 blue ball. Each colour is denoted by its starting letter in capital. The size of the sample space is $|\Omega| = \texttt{colours}^{\texttt{picks}} = 3^2 = 9$, however since there is only one blue ball in the urn, it is not possible to pick a second blue ball, given that one blue ball already has been picked. The possible outcomes are:

$$\Omega = \{\texttt{RR, RO, RB , OO, OR, OB, BR, BO}\}$$

Question 2

Below is a table showing the event of picking two balls from the urn. The first column specifies the given event e.g. the event of picking two red balls. The second column shows the probability of drawing the first colour, this probability is calculated by dividing the number of balls of the given colour with the total amount of balls in the urn. The third column shows the probability of choosing the second ball. Here the probability is calculated as $\mathsf{Pr}[\text{"second draw"}] = \mathsf{Pr}[\text{"second draw"}|\text{"first draw"}]$. The last column shows the probability of the specified event. The probability calculated as $\mathsf{Pr}[\text{"first draw"}] \times \mathsf{Pr}[\text{"second draw"}]$.

event	Pr["First draw"]	Pr["Second draw"]	Pr[event]
RR	$5/9$	$4/8$	$5/9 \times 4/8 = 5/18$
RO	$5/9$	$3/8$	$5/9 \times 3/8 = 5/24$
RB	$5/9$	$1/8$	$5/9 \times 1/8 = 5/72$
OO	$3/9$	$2/8$	$3/9 \times 2/8 = 1/12$
OR	$3/9$	$5/8$	$3/9 \times 5/8 = 5/24$
OB	$3/9$	$1/8$	$3/9 \times 1/8 = 1/24$
BR	$1/9$	$5/8$	$1/9 \times 5/8 = 5/72$
BO	$1/9$	$3/8$	$1/9 \times 3/8 = 1/24$

Question 3

As two balls are picked, and the urn contains 3 orange balls, the the possible values for X is:

$$X \in \{0, 1, 2\}$$

Question 4

Compute $Pr[X = 0]$, that is the probability that the two balls picked are both either "red" or "blue". I will calculate the probability of $X = 0$, by calculating the probability of not picking an orange ball in the first and second draw, that is the complement of those events happening:

$$Pr[\text{"not orange"}] = (1 - (3/9)) \times (1 - (3/8)) = 5/12$$

therefore:

$$Pr[X = 0] = 5/12$$

Question 5

To compute the expected value of X, i will use the general formula:

$$\mathbf{E}[X] = \sum_{x \in \mathbf{X}} x Pr[\mathbf{X} = x] \tag{1}$$

We saw from question 3 that the possible outcomes for X is 0,1 or 2. To get the probability of each possible outcome, each sample space satisfying the values are summed, for the specific value.

$$Pr[X = 0] = \frac{5}{18} + \frac{5}{72} + \frac{5}{72} = \frac{5}{12}$$
$$Pr[X = 1] = \frac{5}{24} + \frac{5}{24} + \frac{1}{24} + \frac{1}{24} = \frac{1}{2}$$
$$Pr[X = 2] = \frac{1}{12}$$

The values can now be used in equation (1):

$$\mathbf{E}[X] = (0 \times \frac{5}{12}) + (1 \times \frac{1}{2}) + (2 \times \frac{1}{12}) = \frac{2}{3}$$

Thus the expected number of orange balls picked is $\frac{2}{3}$

4. Probability Theory: Properties of Expectation

Question 1

Proof of: $\mathbf{E}[X+Y] = \mathbf{E}[X] + \mathbf{E}[Y]$

$$\begin{aligned}
\mathbf{E}[X,Y] &= \sum_{x,y}(x+y)Pr(X=x,Y=y) &(2)\\
&= \sum_{x,y} xPr(X=x,Y=y) + \sum_{x,y} yPr(X=x,Y=y) &(3)\\
&= \sum_{x} x \sum_{y} Pr(X=x,Y=y) + \sum_{y} y \sum_{x} Pr(X=x,Y=y) &(4)\\
&= \sum_{x} xPr(X=x) + \sum_{y} yPr(Y=y) \ \texttt{(By rule of total probability)} &(5)\\
&= \mathbf{E}[X] + \mathbf{E}[Y] &(6)
\end{aligned}$$

Step 3 to 4 uses the rule of total probability : $\sum_x Pr(X=x,Y=y) = Pr(X=x)$

Question 2

Proof of: $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$

$$\begin{aligned}
\mathbf{E}[XY] &= \sum_{x,y} Pr(X=x,Y=y) &(7)\\
&= \sum_{x}\sum_{y} xyPr(X=x,Y=y) &(8)\\
&= \sum_{x}\sum_{y} xyPr(X=x)Pr(Y=y) \ \texttt{(Independent)} &(9)\\
&= \sum_{x} xPr(X=x) \sum_{y} yPr(Y=y) &(10)\\
&= \mathbf{E}[X]\mathbf{E}[Y] &(11)\\
& &(12)
\end{aligned}$$

where step 8 follows from X and Y being independent random variables.

Question 3

k	0	1
0	1/6	3/6
1	1/6	1/6

$$\begin{aligned}
\mathbf{E}[X] &= \left(1 \times \frac{2}{6}\right) + \left(0 \times \frac{4}{6}\right) = \frac{2}{6}\\
\mathbf{E}[Y] &= \left(1 \times \frac{4}{6}\right) + \left(0 \times \frac{2}{6}\right) = \frac{4}{6}\\
\mathbf{E}[XY] &= \left(1 \times 1 \times \frac{1}{6}\right) + \left(1 \times 0 \times \frac{1}{6}\right) + \left(0 \times 1 \times \frac{3}{6}\right) + \left(0 \times 0 \times \frac{1}{6}\right)\\
&= \frac{1}{6}
\end{aligned}$$

Therefore as $\mathbf{E}[X] \times \mathbf{E}[Y] = \frac{1}{2}$ it is evident that $\mathbf{E}[X] \times \mathbf{E}[Y] \neq \mathbf{E}[XY]$

Question 4

Proof of $\mathbf{E}[\mathbf{E}[X]] = \mathbf{E}[X]$
First we write out the inner expected value, thus:

$$\mathbf{E}[\mathbf{E}[X]] = \mathbf{E}\left[\sum_{x \in \mathbf{X}} xPr[\mathbf{X} = x]\right]$$

The inner term no longer contains a random variable, thus the expected value is just the value, therefore

$$\begin{aligned}\mathbf{E}\left[\sum_{x \in \mathbf{X}} xPr[\mathbf{X} = x]\right] &= \sum_{x \in \mathbf{X}} xPr[\mathbf{X} = x] \\ &= \mathbf{E}[X]\end{aligned}$$

Question 5

Proof of $\mathbf{E}\left[(X - \mathbf{E}[X])^2\right] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$:

$$\begin{aligned}\mathbf{E}\left[(X - \mathbf{E}[X])^2\right] &= \mathbf{E}\left[X^2 - 2\mathbf{E}[X]X + (\mathbf{E}[X])^2\right] \\ &= \mathbf{E}[X^2] - 2\mathbf{E}[X]\mathbf{E}[X] + (\mathbf{E}[X])^2 \\ &= \mathbf{E}[X^2] - 2(\mathbf{E}[X])^2 + (\mathbf{E}[X])^2 \\ &= \mathbf{E}[X^2] - (\mathbf{E}[X])^2\end{aligned}$$

Thus proving that $\mathbf{E}\left[(X - \mathbf{E}[X])^2\right] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$.

5. Probability Theory: Complements of Events

Question 1

We denote the complement of an event A by $\overline{A}$. We define the complement as $\overline{A} = \Omega \diagup A$, thus noting that A and its complement are mutually exclusive, meaning we can write $Pr[A \cup B] = Pr[A] + Pr[B]$.
The following is a proof of $Pr[A] = 1 - Pr[\overline{A}]$:

$$\begin{aligned}1 &= Pr[S] = Pr[A \cup \overline{A}] = Pr[A] + Pr[\overline{A}] \\ 1 &= Pr[A] + Pr[\overline{A}] \\ Pr[A] &= 1 - Pr[\overline{A}]\end{aligned}$$

Thus concluding that $Pr[A] = 1 - Pr[\overline{A}]$.

Question 2

If a coin is flipped 10 times, what is the probability of at least one flip being tail. To solve this problem we first denote the aformentioned event as $A =$ `"at least one tail"`, to find the probability of event A i will calculate it's complement which in the above case corresponds to the event $\overline{A} =$ `"all heads"`. The probability of this event can be calculated as followed:

$$\begin{aligned} Pr[A] &= 1 - Pr[\overline{A}] && (13)\\ &= 1 - Pr[\texttt{"all heads"}] && (14)\\ &= 1 - (Pr[H_1] \times Pr[H_2] \times ... \times Pr[H_{10}]) && (15)\\ &= 1 - \left(\frac{1}{2}\right)^{10} \approx 0.999 && (16) \end{aligned}$$

Thus the probability of getting at least one tail is very large.
To calculate the probability of observing at least two tails i will calculate the binomial distribution, which takes parameters n and k, where k is the number of successes in the sequence of n independent experiments. The binomial distribution is calculated as followed:

$$Pr[n,k] = \binom{n}{k} p^k (1-p)^{n-k}$$

For the given example the binomial distribution can be calculated, by taking the complement of the event that either 1 or 0 tails are evident. This is calculated as the following:

$$Pr[A] = 1 - \left(\binom{10}{1}\left(\frac{1}{2}\right)^1\left(1-\frac{1}{2}\right)^9 + \binom{10}{0}\left(\frac{1}{2}\right)^0\left(1-\frac{1}{2}\right)^{10}\right) = \frac{1013}{1024}$$

6. Probability Theory: Coin Flips

A coin is flipped 10 times the following sub assignments are about finding probabilities of different events.

Question 1

As the coin was flipped 10 times, the number of heads must be equal to the number of tails iff. there are 5 heads and 5 tails. Again i will use the binomial distribution to calculate the probability of the event:

$$Pr[10,5] = \binom{10}{5}\left(\frac{1}{2}^5\right)\left(1-\frac{1}{2}\right)^5 \approx 0.246$$

Question 2

The probability of getting more heads than tails will be calculated using binomial distruibtion. The probability must be added for the events of getting 6,7,8,9 and 10 heads, as any of these events will satisfy the question. This probability can be expressed as followed:

$$Pr[\texttt{"more heads than tails"}] = \frac{\binom{10}{6}+\binom{10}{7}+\binom{10}{8}+\binom{10}{9}+\binom{10}{10}}{2^{10}} \approx 0.376$$

Question 3

The *ith* flip and the (11-i)*th* flip are the same for i = 1,2,3,4,5. Flipping the coin for i = 1,2,3,4,5 generates 2^5 possible outcomes. Thus yielding the probability:

$$\frac{2^5}{2^{10}} = \frac{1}{32}$$