Machine Learning Assignment 1

Dhruv Chauhan

November 27, 2016

1 Vectors and Matrices

1.1

$$\mathbf{a} \cdot \mathbf{b} = 3 + 4 + 2 = 9$$

1.2

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

1.3

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

No, this is not equal to 1.1.

1.4

$$\mathbf{b}^T\mathbf{a} = \begin{bmatrix} 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = 9$$

Yes, this is equal to 1.1.

1.5

Using row-reduction (Gauss-Jordan elimination)

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 0 & 0 & 1 \end{array}\right) \xrightarrow{(1/4)\times r_2} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1/4 & 0 \\ 0 & 0 & 2 & 0 & 0 & 1 \end{array}\right) \xrightarrow{(1/2)\times r_3} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1/4 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1/2 \end{array}\right)$$

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

1.6

$$\mathbf{Ma} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 8 \\ 4 \end{bmatrix}$$

1.7

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

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

1.8

The rank of is 1, as $R_2 = 2R_1$, and $R_3 = R_2$.

1.9

If a square matrix is invertible, the number of columns should equal the rank. From this, we can deduce that $\mathbf{A}$ is not invertible as the rank is 1, and therefore not equal to the number of columns, which is 3.

2 Derivatives

2.1

$$\begin{aligned}
f(x) &= \frac{1}{1+e^{-x}} = (1+e^{-x})^{-1} \\
\frac{df(x)}{dx} &= \frac{d}{du}\frac{1}{u} \qquad \text{where } u = 1+e^{-x} \\
u &= 1+e^{-x}, \qquad u' = -e^{-x}
\end{aligned}$$

using the chain rule and substituting back in,

$$\begin{aligned}
\frac{df(x)}{dx} &= -(1+e^{-x})^{-2} \times -e^{-x} \\
= \frac{e^{-x}}{(1+e^{-x})^2} &= \frac{e^{x}}{(1+e^{x})^2}
\end{aligned}$$

2.2

$$\begin{aligned}
\frac{\partial f(w,x)}{\partial w} \qquad & \text{where } f(w,x) = 2(wx+5)^2 \\
&= 4(wx+5) \times x = (4wx+20) \times x \\
&= 4wx^2 + 20x
\end{aligned}$$

3 Probability Theory: Sample Space

3.1

$$\Omega = \{(R,R),(R,O),(R,B),(O,R),(O,O),(O,B),(B,R),(B,O)\}$$

where for (x,y) x is the first ball chosen, y is the second.

3.2

		1^{st} Ball		
		Red	Orange	Blue
	Red	$^{20}/_{72}$	$^{15}/_{72}$	$^{5}/_{72}$
2^{nd} Ball	Orange	$^{15}/_{72}$	$^{6}/_{72}$	$^{3}/_{72}$
	Blue	$^{5}/_{72}$	$^{3}/_{72}$	0

3.3

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

3.4

$$P(X=0) = P(R,R) + P(R,B) + P(B,R) = \frac{30}{72}$$

3.5

$$\mathbb{E}[X] = \sum_{0}^{2} xp(x) = 0 \cdot \frac{30}{72} + 1 \cdot \frac{36}{72} + 2 \cdot \frac{6}{72} = \frac{48}{72} \simeq 0.67$$

4 Probability Theory: Properties of Expectation

4.1

$$\mathbb{E}[X+Y] = \sum_{x,y}(x+y)P(x=X, y=Y) = \sum_{x,y} xP(x=X) + yP(y=Y) =$$
$$\sum_{x} xP(x=X) + \sum_{y} yP(y=Y) = \mathbb{E}[X] + \mathbb{E}[Y]$$

4.2

$$\mathbb{E}[XY] = \sum_{x,y} xy\ P(x=X, y=Y) =$$
$$\sum_{x,y} xy\ P(x=X)P(y=Y) =$$

where P(x = X, y = Y) = P(x = X) P(y = Y) only if X and Y are independent

$$\sum_{x} x\ P(x=X)\ \sum_{y} y\ P(y=Y) = \mathbb{E}[X]\ \mathbb{E}[Y]$$

4.3

Consider the example where a fair die is rolled, and X = 1 if the number is even, $X = \{2, 4, 6\}$ and 0 otherwise. Similarly Y = 1 if the number is prime, $Y = \{2, 3, 5\}$ and 0 otherwise.

The joint distribution table is below:

Y \ X	0	1
0	1/6	4/6
1	2/6	1/6

$$\mathbb{E}[X] = 0 \cdot \frac{3}{6} + 1 \cdot \frac{3}{6} = \frac{1}{2}$$
$$\mathbb{E}[Y] = 0 \cdot \frac{3}{6} + 1 \cdot \frac{3}{6} = \frac{1}{2}$$
$$\therefore \mathbb{E}[X]\ \mathbb{E}[Y] = \frac{1}{4}$$
$$\mathbb{E}[XY] = 0 \cdot \frac{1}{6} + 1 \cdot \frac{1}{6} = \frac{1}{6}$$

4.4

$$\mathbb{E}[X] = \sum_{x} x\ p(x)$$
$$\mathbb{E}[\mathbb{E}[X]] = \sum_{x}(\sum_{x} x\ p(x)) = \sum_{x} x\ p(x)$$
$$= \mathbb{E}[X]$$

4.5

$$
\begin{aligned}
Var[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] &= \mathbb{E}[(X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X])] \\
&= \sum x^2 \; p(x) - 2\mathbb{E}[X] \cdot \sum x \; p(x) + (\mathbb{E}[X])^2 \\
&= \mathbb{E}[X^2] - 2\mathbb{E}[X] \cdot \mathbb{E}[X] + (\mathbb{E}[X])^2 \\
&= \mathbb{E}[X^2] - 2(\mathbb{E}[X])^2 + (\mathbb{E}[X])^2 \\
&= \mathbb{E}[X^2] - (\mathbb{E}[X])^2
\end{aligned}
$$

5 Probability Theory: Complements of Events

5.1

$$\begin{aligned} \overline{A} &= \Omega \setminus A \\ A \cup \overline{A} &= \Omega \\ P(A \cup \overline{A}) &= P(\Omega) = 1 \end{aligned}$$

As by definition we know that A and $\overline{A}$ are mutually exclusive, then we know that

$$\begin{aligned} P(\bigcup_{n=1} E_n) &= \sum_{n=1} P(E_n) \\ P(A) + P(\overline{A}) &= 1 \\ P(A) &= 1 - P(\overline{A}) \end{aligned}$$

5.2

We say that X is the number of tails observed in the 10 flips of a fair coin.

5.2.1

$$\begin{aligned} P(X \geq 1) &= 1 - P(X < 1) \\ P(X < 1) &= P(X = 0) = (\frac{1}{2^{10}}) = \frac{1}{1024} \\ 1 - P(X < 1) &= 1 - \frac{1}{1024} = \frac{1023}{1024} \approx 0.999 \end{aligned}$$

5.2.2

$$\begin{aligned} P(X \geq 2) &= 1 - P(X < 2) \\ P(X < 2) &= P(X = 0) + P(X = 1) \\ P(X = 0) &= \frac{1}{1024} \\ P(X = 1) &= 10 \cdot \frac{1}{1024} = \frac{10}{1024} \\ 1 - P(X < 2) &= 1 - \frac{1}{1024} + \frac{10}{1024} \\ &= \frac{1013}{1024} \approx 0.990 \end{aligned}$$

6 Probability Theory: Coin Flips

As we are flipping a fair coin we say that X is the number of heads, and

$$X \sim B(10, 0.5)$$

6.1

We know that the number of heads and tails can only be equal when X = 5, therefore

$$P(X = 5) = \binom{10}{5} \frac{1}{2^5} \cdot (1 - \frac{1}{2})^5$$
$$= \frac{252}{1024} \approx 0.246$$

6.2

$$P(X > 5) = P(X = 6) + P(X = 7) + ...P(X = 10)$$
$$= \frac{1}{2^{10}} \sum_{i=6}^{10} \binom{10}{i}$$
$$= \frac{386}{1024} \approx 0.317$$

6.3

For the first 5 flips, there are 2^5 possible arrangements of heads and tails, which would be repeated in the inverse for the following 5 flips, so the probability can be calculated as

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