CS 5350/6350: Machine Learining Spring 2019

Homework 0

Handed out: 7 January, 2019
Due: 11:59pm, 16 January, 2019

Author: Cade Parkison
uID: u0939163

Basic Knowledge Review

1. [5 points] We use sets to represent events. For example, toss a fair coin 10 times, and the event can be represented by the set of "Heads" or "Tails" after each tossing. Let a specific event A be "at least one head". Calculate the probability that event A happens, i.e., $p(A)$.

Solution

$$\begin{aligned} p(A) &= p(\text{At least one head}) \\ &= 1 - p(\text{All tails}) \\ &= 1 - \frac{1}{2^{10}} \\ &= \boxed{\frac{1023}{1024}} \end{aligned}$$

2. [10 points] Given two events A and B, prove that

$$p(A \cup B) \leq p(A) + p(B).$$

When does the equality hold?

Solution

Starting with the addition rule of probability:

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

The last term, $p(A \cap B)$, must be greater than or equal to zero. Substituting this into the above equation gives the following:

$$p(A \cup B) \leq p(A) + p(B).$$

3. [10 points] Let $\{A_1, \ldots, A_n\}$ be a collection of events. Show that

$$p(\cup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} p(A_i).$$

When does the equality hold? (Hint: induction)

Solution

Since $p(A \cup B) = p(A) + p(B) - p(A \cap B)$ we have

$$p(\cup_{i=1}^{n+1} A_i) = p(\cup_{i=1}^{n} A_i) + p(A_{n+1}) - p(\cup_{i=1}^{n} A_i \cap A_{n+1})$$

But since

$$p(\cup_{i=1}^{n} A_i \cap A_{n+1}) \geq 0$$

We have

$$p(\cup_{i=1}^{n+1} A_i) \leq p(\cup_{i=1}^{n} A_i) + p(A_{n+1})$$

Therefore

$$p(\cup_{i=1}^{n+1} A_i) \leq \sum_{i=1}^{n+1} p(A_i).$$

In the case where the events A_i are disjoint, the less than equals sign above becomes an equals sign.

4. [20 points] We use $\mathbb{E}(\cdot)$ and $\mathbb{V}(\cdot)$ to denote a random variable's mean (or expectation) and variance, respectively. Given two discrete random variables X and Y, where $X \in \{0,1\}$ and $Y \in \{0,1\}$. The joint probability $p(X,Y)$ is given in as follows:

	$Y=0$	$Y=1$
$X=0$	$1/10$	$2/10$
$X=1$	$3/10$	$4/10$

(a) [10 points] Calculate the following distributions and statistics.

i. the the marginal distributions $p(X)$ and $p(Y)$

ii. the conditional distributions $p(X|Y)$ and $p(Y|X)$

iii. $\mathbb{E}(X)$, $\mathbb{E}(Y)$, $\mathbb{V}(X)$, $\mathbb{V}(Y)$

iv. $\mathbb{E}(Y|X=0)$, $\mathbb{E}(Y|X=1)$, $\mathbb{V}(Y|X=0)$, $\mathbb{V}(Y|X=1)$

v. the covariance between X and Y

(b) [5 points] Are X and Y independent? Why?

(c) [5 points] When X is not assigned a specific value, are $\mathbb{E}(Y|X)$ and $\mathbb{V}(Y|X)$ still constant? Why?

Solution

(a)

i. the the marginal distributions $p(X)$ and $p(Y)$

$$\begin{aligned} p(X_{x=0}) &= p(X_{x=0}, Y_{y=0}) + p(X_{x=0}, Y_{y=1}) \\ &= \frac{1}{10} + \frac{2}{10} = \frac{3}{10} \end{aligned}$$

$$\begin{aligned} p(X_{x=1}) &= p(X_{x=1}, Y_{y=0}) + p(X_{x=1}, Y_{y=1}) \\ &= \frac{3}{10} + \frac{4}{10} = \frac{7}{10} \end{aligned}$$

$$\begin{aligned} p(Y_{y=0}) &= p(X_{x=0}, Y_{y=0}) + p(X_{x=1}, Y_{y=0}) \\ &= \frac{1}{10} + \frac{3}{10} = \frac{2}{5} \end{aligned}$$

$$\begin{aligned} p(Y_{y=1}) &= p(X_{x=0}, Y_{y=1}) + p(X_{x=1}, Y_{y=1}) \\ &= \frac{2}{10} + \frac{4}{10} = \frac{3}{5} \end{aligned}$$

ii. the conditional distributions $p(X|Y)$ and $p(Y|X)$

Using the following formula, we can easily calculate the conditional distributions

$$p(X|Y) = \frac{p(X,Y)}{p(Y)}$$

$p(X|Y)$:

$$\begin{aligned}
p(X_{x=0}|Y_{y=0}) &= \frac{p(X_{x=0}, Y_{y=0})}{\sum_i p(X=i, Y=0)} \\
&= \frac{1/10}{1/10 + 3/10} \\
&= \boxed{\frac{1}{4}}
\end{aligned}$$

$$\begin{aligned}
p(X_{x=1}|Y_{y=0}) &= \frac{p(X_{x=1}, Y_{y=0})}{\sum_i p(X=i, Y=0)} \\
&= \frac{3/10}{1/10 + 3/10} \\
&= \boxed{\frac{3}{4}}
\end{aligned}$$

$$\begin{aligned}
p(X_{x=0}|Y_{y=1}) &= \frac{p(X_{x=0}, Y_{y=1})}{\sum_i p(X=i, Y=1)} \\
&= \frac{2/10}{2/10 + 4/10} \\
&= \boxed{\frac{1}{3}}
\end{aligned}$$

$$\begin{aligned}
p(X_{x=1}|Y_{y=1}) &= \frac{p(X_{x=1}, Y_{y=1})}{\sum_i p(X=i, Y=1)} \\
&= \frac{4/10}{2/10 + 4/10} \\
&= \boxed{\frac{2}{3}}
\end{aligned}$$

$p(Y|X)$:

$$\begin{aligned} p(Y_{y=0}|X_{x=0}) &= \frac{p(X_{x=0}, Y_{y=0})}{\sum_i p(X=0, Y=i)} \\ &= \frac{1/10}{1/10 + 2/10} \\ &= \boxed{\frac{1}{3}} \end{aligned}$$

$$\begin{aligned} p(Y_{y=1}|X_{x=0}) &= \frac{p(X_{x=0}, Y_{y=1})}{\sum_i p(X=0, Y=i)} \\ &= \frac{2/10}{1/10 + 2/10} \\ &= \boxed{\frac{2}{3}} \end{aligned}$$

$$\begin{aligned} p(Y_{y=0}|X_{x=1}) &= \frac{p(X_{x=1}, Y_{y=0})}{\sum_i p(X=1, Y=i)} \\ &= \frac{3/10}{3/10 + 4/10} \\ &= \boxed{\frac{3}{7}} \end{aligned}$$

$$\begin{aligned} p(Y_{y=1}|X_{x=1}) &= \frac{p(X_{x=1}, Y_{y=1})}{\sum_i p(X=1, Y=i)} \\ &= \frac{4/10}{3/10 + 4/10} \\ &= \boxed{\frac{4}{7}} \end{aligned}$$

iii. $\mathbb{E}(X)$, $\mathbb{E}(Y)$, $\mathbb{V}(X)$, $\mathbb{V}(Y)$

$$\begin{aligned} \mathbb{E}(X) &= \sum_x xp(x) \\ &= (0) * p(x=0) + (1) * p(x=1) \\ &= 0 * \frac{3}{10} + 1 * \frac{7}{10} \\ &= \boxed{\frac{7}{10}} \end{aligned}$$

$$
\begin{aligned}
\mathbb{E}(Y) &= \sum_y y p(y) \\
&= (0) * p(y=0) + (1) * p(y=1) \\
&= 0 + \frac{3}{5} \\
&= \boxed{\frac{3}{5}}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{V}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\
&= \sum_x x^2 p(x) - \mathbb{E}(X)^2 \\
&= \frac{7}{10} - (\frac{7}{10})^2 \\
&= \boxed{\frac{21}{100}} = 0.21
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{V}(Y) &= \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 \\
&= \sum_y y^2 p(y) - \mathbb{E}(Y)^2 \\
&= \frac{3}{5} - (\frac{3}{5})^2 \\
&= \boxed{\frac{6}{25}} = 0.24
\end{aligned}
$$

iv. $\mathbb{E}(Y|X=0)$, $\mathbb{E}(Y|X=1)$, $\mathbb{V}(Y|X=0)$, $\mathbb{V}(Y|X=1)$

$$
\begin{aligned}
\mathbb{E}(Y|X=0) &= \sum_y y P(y|x=0) \\
&= (0) * P(y=0|x=0) + (1) * P(y=1|x=0) \\
&= \boxed{\frac{2}{3}}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}(Y|X=1) &= \sum_y yP(y|x=1) \\
&= (0) * P(y=0|x=1) + (1) * P(y=1|x=1) \\
&= \boxed{\frac{4}{7}}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{V}(Y|X=0) &= \sum_y y^2 P(y|X=0) - \mathbb{E}(Y|X=0)^2 \\
&= (0)^2 * P(Y=0|X=0) + (1)^2 * P(Y=1|X=0) - \mathbb{E}(Y|X=0)^2 \\
&= \frac{2}{3} - (\frac{2}{3})^2 \\
&= \boxed{\frac{2}{9}}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{V}(Y|X=1) &= \sum_y y^2 P(y|X=1) - \mathbb{E}(Y|X=1)^2 \\
&= (0)^2 * P(Y=0|X=1) + (1)^2 * P(Y=1|X=1) - \mathbb{E}(Y|X=1)^2 \\
&= \frac{4}{7} - (\frac{4}{7})^2 \\
&= \boxed{\frac{12}{49}} \approx 0.2449
\end{aligned}
$$

v. the covariance between X and Y

$$
\begin{aligned}
Cov(X,Y) &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \\
&= \sum_x \sum_y xyP(x,y) - \mathbb{E}(X)\mathbb{E}(Y) \\
&= \Big((0 \times 0) \times \frac{1}{10} + (0 \times 1) \times \frac{2}{10} + (1 \times 0) \times \frac{3}{10} + (1 \times 1) \times \frac{4}{10}\Big) - \frac{7}{10} \times \frac{3}{5} \\
&= \frac{4}{10} - \frac{21}{50} \\
&= \boxed{-\frac{1}{50}}
\end{aligned}
$$

(b) [5 points] Are X and Y independent? Why?

X and Y are not independent because $Cov(X,Y) \neq 0$. If they were independent, then the following must also hold:

$$\mathbb{E}(Y|X = x) = \mathbb{E}(Y)$$

But, for example, when we plug in $x = 1$ this condition is not met:

$$\mathbb{E}(Y|X = 1) = \mathbb{E}(Y)$$
$$\frac{4}{7} \neq \frac{3}{5}$$

(c) [5 points] When X is not assigned a specific value, are $\mathbb{E}(Y|X)$ and $\mathbb{V}(Y|X)$ still constant? Why?

Yes, these both remain constant. If X is not assigned a specific value, it does not change the possible values that X can obtain. This means that the Expectation and Variance would still be the same.

5. [10 points] Assume a random variable X follows a standard normal distribution, i.e., $X \sim \mathcal{N}(X|0,1)$. Let $Y = e^X$. Calculate the mean and variance of Y.

Solution

(a) $\mathbb{E}(Y)$

The probability density function of a standard normal distribution is:

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Plugging this into the definition of Expected value gives:

$$\begin{aligned}
\mathbb{E}(Y) &= \int_{-\infty}^{\infty} e^x \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2-x} dx \\
&= \boxed{0}
\end{aligned}$$

(b) $\mathbb{V}(Y)$

$$\begin{aligned}
\mathbb{V}(Y) &= \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 \\
&= \int_{-\infty}^{\infty} e^{2x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx - (0)^2 \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2-2x} dx - (0)^2 \\
&= \boxed{0}
\end{aligned}$$

6. [20 points] Given two random variables X and Y, show that

(a) $\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(Y)$

Solution

$$
\begin{aligned}
\mathbb{E}(\mathbb{E}(Y|X)) &= \int_{-\infty}^{\infty} \mathbb{E}(Y|X=x)P(x)dx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yP(y|x)dyP(x)dx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yP(y|x)P(x)dydx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yP(y,x)dydx \\
&= \int_{-\infty}^{\infty} y\int_{-\infty}^{\infty} P(y,x)dydx \\
&= \int_{-\infty}^{\infty} yP(y)dy \\
&= \boxed{\mathbb{E}(Y)}
\end{aligned}
$$

(b) $\mathbb{V}(Y) = \mathbb{E}(\mathbb{V}(Y|X)) + \mathbb{V}(\mathbb{E}(Y|X))$

Solution

Using the definition of expected value:

$$\mathbb{E}(\mathbb{V}(Y|X)) = \mathbb{E}(\mathbb{E}(Y^2|X)) - \mathbb{E}((\mathbb{E}(Y|X))^2)$$

But, from part (a), we can say the following:

$$\mathbb{E}(\mathbb{E}(Y^2|X)) = \mathbb{E}(Y^2)$$

Therefore:

$$\mathbb{E}(\mathbb{V}(Y|X)) = \mathbb{E}(Y^2) - \mathbb{E}((\mathbb{E}(Y|X))^2)$$

Using the other formula for variance, we have:

$$
\begin{aligned}
\mathbb{V}(\mathbb{E}(Y|X)) &= \mathbb{E}(\mathbb{E}(Y|X)^2) - (\mathbb{E}(\mathbb{E}(Y|X))^2 \\
&= \mathbb{E}(\mathbb{E}(Y|X)^2) - (\mathbb{E}(Y))^2
\end{aligned}
$$

Combining the two equations above gives the wanted result:

$$
\begin{aligned}
\mathbb{E}(\mathbb{V}(Y|X)) + \mathbb{V}(\mathbb{E}(Y|X)) &= \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2 \\
&= \mathbb{V}(Y)
\end{aligned}
$$

7. [15 points] Given a logistic function, $f(\mathbf{x}) = 1/(1 + \exp(-\mathbf{a}^\top \mathbf{x}))$ ($\mathbf{x}$ is a vector), derive/calculate the following gradients and Hessian matrices.

Solution

(a) $\nabla f(\mathbf{x})$

$$
\begin{aligned}
\nabla f(\mathbf{x}) &= \frac{\partial}{\partial \mathbf{x}}\Big(\frac{1}{1+e^{-\mathbf{a}^\top\mathbf{x}}}\Big) \\
&= \frac{e^{-\mathbf{a}^\top\mathbf{x}}}{(1+e^{-\mathbf{a}^\top\mathbf{x}})^2} \\
&= \frac{1+e^{-\mathbf{a}^\top\mathbf{x}}-1}{(1+e^{-\mathbf{a}^\top\mathbf{x}})^2} \\
&= \frac{1+e^{-\mathbf{a}^\top\mathbf{x}}}{(1+e^{-\mathbf{a}^\top\mathbf{x}})^2} - \Big(\frac{1}{1+e^{-\mathbf{a}^\top\mathbf{x}}}\Big)^2 \\
&= \frac{1}{(1+e^{-\mathbf{a}^\top\mathbf{x}})} - \Big(\frac{1}{1+e^{-\mathbf{a}^\top\mathbf{x}}}\Big)^2 \\
&= f(\mathbf{x}) - f(\mathbf{x})^2 \\
&= \boxed{f(\mathbf{x})(1-f(\mathbf{x}))}
\end{aligned}
$$

(b) $\nabla^2 f(\mathbf{x})$

$$
\begin{aligned}
\nabla^2 f(\mathbf{x}) &= \frac{\partial}{\partial \mathbf{x}}\big(f(\mathbf{x})(1-f(\mathbf{x}))\big) \\
&= \nabla f(\mathbf{x})(1-f(\mathbf{x})) - f(\mathbf{x})\nabla(f\mathbf{x}) \\
&= \nabla f(\mathbf{x})(1-2f(\mathbf{x})) \\
&= f(\mathbf{x})(1-f(\mathbf{x})(1-2f(\mathbf{x})) \\
&= \boxed{f(\mathbf{x})(2f(\mathbf{x})^2 - 3f(\mathbf{x}) + 1)}
\end{aligned}
$$

(c) $\nabla f(\mathbf{x})$ when $\mathbf{a} = [1,1,1,1,1]^\top$ and $\mathbf{x} = [0,0,0,0,0]^\top$

Say $\mathbf{z} = -\mathbf{a}^\top\mathbf{x}$, plugging in $\mathbf{a}$ and $\mathbf{x}$ gives:

$$
\begin{aligned}
\mathbf{z} &= -\mathbf{a}^\top\mathbf{x} \\
&= 0
\end{aligned}
$$

Plugging this in to part (a) gives the following:

$$
\begin{aligned}
\nabla f(\mathbf{x}) &= \frac{1}{(1+e^0)}\Big(1 - \frac{1}{(1+e^0)}\Big) \\
&= \frac{1}{2}(1 - \frac{1}{2}) \\
&= \boxed{-\frac{1}{4}}
\end{aligned}
$$

(d) $\nabla^2 f(\mathbf{x})$ when $\mathbf{a} = [1,1,1,1,1]^\top$ and $\mathbf{x} = [0,0,0,0,0]^\top$
Evaluating $f(\mathbf{x})$ at these values gives:

$$
\begin{aligned}
f(\mathbf{x}) &= 1/(1+\exp(-z)) \\
&= 1/(1+\exp(0)) \\
&= \frac{1}{2}
\end{aligned}
$$

Plugging in this value to the $\nabla^2 f(\mathbf{x})$ equation gives the following:

$$
\begin{aligned}
\nabla^2 f(\mathbf{x}) &= f(\mathbf{x})(2f(\mathbf{x})^2 - 3f(\mathbf{x}) + 1) \\
&= (\frac{1}{2}) \times \Big(2 \times (\frac{1}{2})^2 - 3 \times \frac{1}{2} + 1\Big) \\
&= \boxed{0}
\end{aligned}
$$

Note that $0 \le f(\mathbf{x}) \le 1$.

8. [10 points] Show that $g(x) = -\log(f(\mathbf{x}))$ where $f(\mathbf{x})$ is a logistic function defined as above, is convex.

Solution

We can prove the convexity of $g(\mathbf{x})$ by applying the second-order condition of convexity. This states that a twice-differentiable function is convex if and only if its hessian matrix is positive semi-definite. Given the definition of the logistic function and the equation of the hessian, we can conclude that the hessian is indeed positive semi-definite since the logistic function is always non-negative. Therefore, $g(\mathbf{x})$ is indeed convex.