CS 5350/6350: Machine Learining Spring 2019

Homework 0

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

- You are welcome to talk to other members of the class about the homework. I am more concerned that you understand the underlying concepts. However, you should write down your own solution. Please keep the class collaboration policy in mind.
- Feel free discuss the homework with the instructor or the TAs.
- Your written solutions should be brief and clear. You need to show your work, not just the final answer, but you do *not* need to write it in gory detail. Your assignment should be **no more than 10 pages**. Every extra page will cost a point.
- Handwritten solutions will not be accepted.
- The homework is due by **midnight of the due date**. Please submit the homework on Canvas.
- Some questions are marked **For 6350 students**. Students who are registered for CS 6350 should do these questions. Of course, if you are registered for CS 5350, you are welcome to do the question too, but you will not get any credit for it.

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., $p(A)$.
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?
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)
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:

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

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

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

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?

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.

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

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

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

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

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

   (Hints: using definition.)

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.

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

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

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

   (d) $\nabla^2 f(\mathbf{x})$ when $\mathbf{a} = [1,1,1,1,1]^\top$ and $\mathbf{x} = [0,0,0,0,0]^\top$

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

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

Answers:

1. We consider this question with contradiction, which means to find the probability with no "Heads" (TTTTTTTTTT) = 1. Total number of case is 2ˆ10 = 1024. Then we get p(A):

$$p(A) = 1 - 1/1024 = 1023/1024$$

2. From the formula:

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

   When

$$p(A \cap B) = 0, p(A \cup B) = p(A) + p(B).$$

When

$$0 < p(A \cap B) \leq 1, p(A \cup B) + p(A \cap B) = p(A) + p(B)$$
$$p(A \cup B) < p(A) + p(B)$$

Therefore,

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

3. We prove by induction:
For the n = 1 case, we have

$$p(A_1) \leq p(A_1)$$

Suppose we have case n with

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

Since

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

we have

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

Since

$$p(\cup_{i=1}^{n} A_i \cap A_{n+1}) > 0,$$
$$p(\cup_{i=1}^{n+1} A_i) \leq p(\cup_{i=1}^{n} A_i) + p(A_{n+1})$$
$$p(\cup_{i=1}^{n+1} A_i) \leq \sum_{i=1}^{n} p(A_i) + p(A_{n+1}) = \sum_{i=1}^{n+1} p(A_i)$$

Therefore,

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