

CS 475 Machine Learning: Homework 1

Learning Foundations

Due: Monday September 14, 2015, 11:59pm

50 Points Total Version 1.0

Li-Yi Lin / llin34@jhu.edu

1 Analytical (15 Points)

1 (4 points) Suppose you are given a biased coin. Specifically, if you flip the coin, it comes up heads with probability $5/6$, and comes up tails with probability $1/6$. What is the expected number of flips such that the coin comes up tails exactly once.

Ans: Since we want to know the expected number of flips that there is exactly one tail happen in those flips, the tail must happen in the last flip and we stop toss the coin. So we can calculate the expected number by

$$\sum_{x=1}^{\infty} x \times (5/6)^{x-1} \times 1/6 = 6$$

x is the number of flips. So the expected number of flips that the coin comes up tails exactly once is 6.

2 (3 points) Suppose a family have two children. Given the fact that one is a girl, what is the probably that the other child is also a girl?

Ans: Based one the given problem description, we can only assume that the only factor that determines the probability of a baby being girl or boy is the Y-chromosome comes from the baby's father (woman only have X-chromosome). If a Y-chromosome from a man combines with a X-chromosome from a woman, then the baby will be a boy. On the other hand, if a X-chromosome form a man combines with a X-chromosome from a woman, then the baby will be a girl. Then, we further assume that the probabilities of X-chromosome or Y-chromosome comes from a man are both $1/2$ and are independent and identically distributed. Therefore, the probability of the other child is also a girl is still $1/2$.

3 (4 points) Please show that the following loss functions are convex: (a) $\ell(y, \hat{y}) = |y - \hat{y}|$; (b) $\ell(y, \hat{y}) = (y - \hat{y})^2$.

Ana:

(a) Assume \hat{y} is fixed. $\forall y_1, y_2 \in R, \forall t \in [0, 1]$ if we can prove that $\ell(t \times y_1 + (1-t) \times y_2, \hat{y}) \leq t \times \ell(y_1, \hat{y}) + (1-t) \times \ell(y_2, \hat{y})$, then the loss function is a convex. The above equation can be written as:

$$\ell(t \times y_1 + (1-t) \times y_2, \hat{y}) \leq t \times \ell(y_1, \hat{y}) + (1-t) \times \ell(y_2, \hat{y})$$

We start from the left-hand side of the equation:

$$\ell(t \times y_1 + (1-t) \times y_2, \hat{y}) = |t \times y_1 + (1-t) \times y_2 - \hat{y}|$$

$$\begin{aligned}
&= |t \times y_1 + (1 - t) \times y_2 - t \times \hat{y} - (1 - t) \times \hat{y}| \\
&\leq |t \times y_1 - t \times \hat{y}| + |(1 - t) \times y_2 - (1 - t) \times \hat{y}| \\
&= t \times |y_1 - \hat{y}| + (1 - t) \times |y_2 - \hat{y}| \\
&= t \times \ell(y_1, \hat{y}) + (1 - t) \times \ell(y_2, \hat{y})
\end{aligned}$$

Since we have proved that $\ell(t \times y_1 + (1 - t) \times y_2, \hat{y}) \leq t \times \ell(y_1, \hat{y}) + (1 - t) \times \ell(y_2, \hat{y})$, the loss function is convex.

(b) To prove the loss function, $\ell(y, \hat{y}) = (y - \hat{y})^2$, is convex, we will need to prove the second derivative of the equation on y is a non-negative function. The second derivative of the loss function is

$$\frac{\partial^2 \ell(y, \hat{y})}{\partial y^2} = 2 > 0$$

Since the second derivative is 2 and it is always large than 0, the loss function is convex.

4 (3 points) Give an example of an optimal hypothesis, a finite hypothesis class that contains the optimal hypothesis, and an infinite class that does not contain the optimal hypothesis.

Ans: Assume we have a fair dice. The optimal hypothesis of the probability of getting a "1" from a toss of this dice is $1/6$. A finite hypothesis class might be $\{1/6, 2/6, 3/6\}$. A infinite hypothesis class might be those probabilities other than that in the finite hypothesis class, i.e. any number in $[0,1] - \{1/6, 2/6, 3/6\}$.

5 (1 point) How are you comfortable with probability and linear algebra?

Ans: I had taken probability and linear algebra courses before and also reviewed them during this summary. I am comfortable with using them on the Machine Learning course.