

# 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 tossing the coin. So we can calculate the expected number by

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

$x$  is the number of flips. To solve this summation, we transform it to the new form shown below

$$\begin{aligned} & \sum_{x=0}^{\infty} \{(x+1) \times (5/6)^x \times 1/6\} \\ &= 1/6 \times \left\{ \sum_{x=0}^{\infty} x \times (5/6)^x + \sum_{x=0}^{\infty} (5/6)^x \right\} \\ &= 1/6 \times \left\{ \sum_{x=0}^{\infty} x \times (5/6)^x + 6 \right\} \\ &= \sum_{x=0}^{\infty} \{x \times (5/6)^x \times 1/6\} + 1 \\ &= 5/6 \times \sum_{x=1}^{\infty} \{x \times (5/6)^{x-1}\} + 1 \end{aligned}$$

Then we got

$$\begin{aligned} E(x) &= 5/6 \times E(x) + 1 \\ E(x) &= 6 \end{aligned}$$

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 probability that the other child is also a girl?

**Ans:** Assume the probabilities of a child being girl or boy are both  $1/2$  and the probability are independent and identically distributed. Use  $G$  denotes a girl and  $B$  denotes a boy. So the probability that the other child is also a girl (event A) given that we know one child is a girl (event B) can be calculated as following:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Since the probability of event A and B are independent, the above equation becomes

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \times P(B)}{P(B)} = P(A) = 1/2$$

Therefore, the probability that the other child is also a girl given one child is a girl is  $1/2$ .

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

**Ans:**

(a) Assume  $\hat{y}$ , the true value of  $y$ , is fixed for each instance.  $\forall y_1, y_2 \in R, \forall t \in [0, 1]$  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. We start from the left-hand side of the equation:

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

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 can be  $\{1/6, 2/6, 3/6\}$ . A infinite hypothesis class can be 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 learned probability and linear algebra courses when I was a undergraduate student and also reviewed them during this summer. I am comfortable with using them on the Machine Learning course.