Homework 3

October 20, 2016

Maximum Likelihood Estimation

Exercise 1. Suppose we want to investigate the number of students arriving at ShanghaiTech Dining Hall per minute during noon time. The sample data collected yesterday between 11:00am-1:00pm is

Number of Students	0	1	2	3	4
Occurrence	40	281	134	19	6

Using MLE, find the expected number of students arriving the Dining Hall per minute.

Perceptron

Exercise 2. The weight update rule in Perceptron Learning Algorithm (PLA) has the nice interpretation that it moves in the direction of classifying $x(t)$ correctly at iteration t.

1. Show that $y(t)w^T(t)x(t) \leq 0$. [Hint: $x(t)$ is misclassified by $w(t)$]
2. Show that $y(t)w^T(t+1)x(t) > y(t)w^T(t)x(t)$.
3. As far as classifying $x(t)$ is concerned, argue that the move from $w(t)$ to $w(t+1)$ is a move "in the right direction".

Hoeffding Inequality

Exercise 3. 1. Using binomial distribution, find the probability that a sample of 10 marbles will have $\nu \leq 0.1$ given that $\mu = 0.9$.

2. Using the Hoeffding Inequality to bound the probability that a sample of 10 marbles will have $\nu \leq 0.1$ given that $\mu = 0.9$.

Learning feasibility

Exercise 4. We are given a data set $\mathcal{D}$ of 25 training examples from an unknown target function $f : \mathcal{X} \to \mathcal{Y}$ where $\mathcal{X} = \mathbb{R}$ and $\mathcal{Y} = \{-1, +1\}$. To learn f, we use a simple hypothesis set $\mathcal{H} = \{h_1, h_2\}$ where h_1 is the constant $+1$ function and h_2 is the constant -1.

We consider two learning algorithms, S (smart) and C (crazy). S chooses the hypothesis that agrees the most with $\mathcal{D}$ and C chooses the other hypothesis deliberately. Let us see how these algorithms perform out of sample from the deterministic and probabilistic points of view. Assume in the probabilistic view that there is a probability distribution on $\mathcal{X}$, and let $\mathbb{P}[f(\mathbf{X}) = +1] = p$.

1. Can S produce a hypothesis t hat is guaranteed to perform better than random on any point outside $\mathcal{D}$?

2. Assume for the rest of the exercise that all the examples in $\mathcal{D}$ have $y_n = +1$. Is it possible that the hypothesis that C produces turns out to be better than the hypothesis that S produces?

3. If $p = 0.9$, what is the probability that S will produce a better hypothesis that C?

4. Is there any value of p for which it is more likely than not that C will produce a better hypothesis than S?