CS221 Fall 2017 Homework [Foundations]

Name: yf-yang

By turning in this assignment, I agree by the Stanford honor code and declare that all of this is my own work.

Problem 1

(a) Let $\theta^* = arg\min_\theta f(\theta)$, then

$$\frac{\partial f(\theta^*)}{\partial \theta^*} = \sum_{i=1}^{n} w_i(\theta^* - x_i) = 0 \tag{1}$$

$$\theta^* \sum_{i=1}^{n} w_i = \sum_{i=1}^{n} w_i x_i \tag{2}$$

$$\theta^* = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \tag{3}$$

Without loss of generality, we can assume $\sum_{i=1}^{n} w_i = 1$, then

$$\theta^* = \mathbb{E}_{\mathbf{w}}[\mathbf{x}] \tag{4}$$

In order to prove θ^* is a global minimum, we need to show

$$\frac{\partial^2 f(\theta^*)}{\partial {\theta^*}^2} = \sum_{i=1}^{n} w_i \tag{5}$$

Note that Eqn(5) is strictly positive if all w_i are positive, so θ^* should be a global minimum, while when some w_i is negative, $sum_{i=1}^{n} w_i$ may be negative, then the global minimum doesn't hold. $\square$

(b)

$$f(\mathbf{x}) = \sum_{i=1}^{d} \max_{s \in \{-1,1\}} s x_i = \sum_{i=1}^{d} |x_i| \tag{6}$$

$$g(\mathbf{x}) = \max_{s \in \{-1,1\}} \sum_{i=1}^{d} s x_i = \max_{s \in \{-1,1\}} s \sum_{i=1}^{d} x_i = |\sum_{i=1}^{d} x_i| \tag{7}$$

According to Triangular Inequality, we can conclude that $f(\mathbf{x}) > g(\mathbf{x})$ $\square$

(c) Let $\mathbf{X_i}$ = (number of points you roll the i-th time), $\mathbf{N}$ = (number of rolls), then

$$\begin{aligned}
\mathbb{E}[\sum_{i=1}^{\mathbf{N}} \mathbf{X_i}] &= \sum_{n=0}^{\infty} f_{\mathbf{N}}(n+1)\,\mathbb{E}[\mathbf{X}|n+1] \\
&= \sum_{n=0}^{\infty} \frac{1}{6}(\frac{5}{6})^n \frac{b-a}{5} n \\
&= \frac{b-a}{30} \sum_{n=0}^{\infty} n(\frac{5}{6})^n \\
&= b-a
\end{aligned}$$

□

(d) Let $M(p) = \log L(p)$, then

$$M(p) = 4\log p + 3\log(1-p) \tag{8}$$

At $M(p)$'s maximum, which is also $L(p)$'s maximum we have

$$\frac{\partial M(p)}{\partial p} = \frac{4}{p} - \frac{3}{1-p} = 0 \tag{9}$$

$$p = \frac{4}{7} \tag{10}$$

Intuitively, in order to get 4 Heads in 7 flips with max probability(on average $\frac{4}{7}$ Head in one flip), we should let the probability of Heads in one flip, which is the Expectation of Heads in one flip be exactly $\frac{4}{7}$(because $L(p)$ is a continuous function and Expectation of Heads in one flip should be where the probability is the maximum. □

(e)

$$\frac{\partial f}{\partial w_k} = \sum_{i=1}^{n}\sum_{j=1}^{n} 2(a_{ik} - b_{jk})(\mathbf{a_i}^\top \mathbf{w} - \mathbf{b_j}^\top \mathbf{w}) + 2\lambda w_k \tag{11}$$

$$\frac{\partial f}{\partial \mathbf{w}} = 2(\sum_{i=1}^{n}\sum_{j=1}^{n} (\mathbf{a_i}^\top \mathbf{w} - \mathbf{b_j}^\top \mathbf{w})(\mathbf{a_i} - \mathbf{b_j}) + \lambda \mathbf{w}) \tag{12}$$

□

Problem 2

(a) Start from 1-dimensional case with only one line segment to be placed, then the number of cases is

$$\sum_{i=1}^{n} n = \frac{n(n+1)}{2} \tag{13}$$

With 6 independent parts and two independent axis, the total number should be

$$((\frac{n(n+1)}{2})^6)^2 = \frac{n(n+1)}{2})^{12} \tag{14}$$

And

$$\lim_{n\to\infty} \frac{n(n+1)}{2})^{12} = O(n^{24}) \tag{15}$$

□

(b) See prob2-2.py for the code. The complexity is $O(n^2)$. □

(c)

$$f(n) = \begin{cases} 1, n = 0, 1 \\ \sum_{i=0}^{n-1} f(i), n \geq 2 \end{cases} \tag{16}$$

Then, we can prove $f(n)$ has a more direct form:

$$f(n) = 2^{n-1}, n \in \mathbb{N}^+ \tag{17}$$

We can prove that by induction:
Base Case: Obviously, we have $f(1) = 2^0 = 1$
Inductive Step:

$$f(n) = \sum_{i=1}^{n-1} 2^{i-1} + f(0) = 2^{n-1} - 1 + 1 = 2^{n-1} \tag{18}$$

□

(d)

$$\sum_{i=1}^{n}\sum_{j=1}^{n} (\mathbf{a_i}^\top \mathbf{w} - \mathbf{b_j}^\top \mathbf{w})^2 + \lambda \left\|\mathbf{w}\right\|_2^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} (\mathbf{w}^\top \mathbf{a_i}\mathbf{a_i}^\top \mathbf{w} - \mathbf{w}^\top \mathbf{a_i}\mathbf{b_j}^\top \mathbf{w} + \mathbf{w}^\top \mathbf{b_j}\mathbf{b_j}^\top \mathbf{w}) + \lambda \left\|\mathbf{w}\right\|_2^2 \tag{19}$$

$$= \mathbf{w}^\top (\sum_{i=1}^{n}\sum_{j=1}^{n} (\mathbf{a_i}\mathbf{a_i}^\top - \mathbf{a_i}\mathbf{b_j}^\top + \mathbf{b_j}\mathbf{b_j}))\mathbf{w} + \lambda \left\|\mathbf{w}\right\|_2^2 \tag{20}$$

See prob2-4.py for the code. □