

Homework 1

Prabhjot Singh Rai

September 30, 2018

1 Optimization and probability

Solution 1a: In order to check whether the minima exists for the function, $f(\theta)$, we calculate $f''(\theta)$.

$$\begin{aligned}\frac{df(\theta)}{d\theta} &= \sum_{i=1}^n w_i(\theta - x_i) \\ g(\theta) &= \sum_{i=1}^n w_i\theta - \sum_{i=1}^n w_ix_i\end{aligned}$$

Differentiating $g(\theta)$, we get

$$\frac{dg(\theta)}{d\theta} = \sum_{i=1}^n w_i \tag{1}$$

Since it is given that w_1, \dots, w_n are positive real numbers, therefore $f''(\theta) > 0$, which shows that the graph of this quadratic function is concave up and minima exists at value of θ when $f'(\theta) = 0$.

$$\begin{aligned}\frac{df(\theta)}{d\theta} &= \sum_{i=1}^n w_i(\theta - x_i) \\ 0 &= \sum_{i=1}^n w_i\theta - \sum_{i=1}^n w_ix_i \\ \theta &= \frac{\sum_{i=1}^n w_ix_i}{\sum_{i=1}^n w_i}\end{aligned}$$

Therefore, the value of θ which minimizes $f(\theta)$ is $\frac{\sum_{i=1}^n w_ix_i}{\sum_{i=1}^n w_i}$. If some of the w_i 's are negative, then from equation (1), it's not necessary that $f''(\theta)$ is

always positive. If $f''(\theta) < 0$, the function may not have any defined minima, since it will be concave downwards. And if a point of inflection exists ($f''(\theta) = 0$), then again, there would be local minima and maxima but no defined global minima for the function.

Solution 1b. Let n vectors out of x_1, \dots, x_d be negative where $n \leq d$. Let sum of negative vectors be $-\beta$ and sum of positive vectors be α , where α and β are both positive vectors. Evaluating $f(x)$,

$$\begin{aligned} f(x) &= \sum_{i=1}^d \max_{s \in \{1, -1\}} s x_i \\ &= \max_{s \in \{1, -1\}} s(\alpha) + \max_{s \in \{1, -1\}} s(-\beta) \\ &= \alpha + \beta \end{aligned}$$

Evaluating $g(x)$,

$$\begin{aligned} g(x) &= \max_{s \in \{1, -1\}} \sum_{i=1}^d s x_i \\ &= \max_{s \in \{1, -1\}} s \sum_{i=1}^d x_i \\ &= \max_{s \in \{1, -1\}} s(\alpha - \beta) \\ g(x) &= \begin{cases} \alpha - \beta & \text{for } \alpha > \beta \\ \beta - \alpha & \text{for } \beta > \alpha \\ 0 & \text{for } \alpha = \beta \end{cases} \end{aligned}$$

Comparing $f(x)$ and $g(x)$ for different cases:

$$f(x) - g(x) = \begin{cases} 2\beta & \text{for } \alpha > \beta \\ 2\alpha & \text{for } \beta > \alpha \\ 2\alpha = 2\beta & \text{for } \alpha = \beta \end{cases}$$

Therefore, if α and β are non-zero, $f(x) > g(x)$, else $f(x) = g(x)$. Hence, for all \mathbf{x} , $f(x) \geq g(x)$.

Solution 1c. Let $E(a, b)$ denote the expected number of points when we stop. Only cases to consider would be 1, 2 and 6. Let's divide the sequences into 3 cases.

One third times we get a 1 first, and then we stop.
 One third times we get a 2 first, and then we repeat.
 One third times we get a 6 first, and then we repeat.

Expressing in equation form, it would be

$$\begin{aligned} E(x) &= \frac{1}{3}(0) + \frac{1}{3}(-a + E(x)) + \frac{1}{3}(b + E(x)) \\ 3E(x) &= -a + b + 2E(x) \\ E(x) &= b - a \end{aligned}$$

Solution 1d. Since we know that value of p for which $\log L(p) = 0$ must also satisfy $L(p) = 0$, checking the point where the slope of the function is zero

$$\begin{aligned} \log L(p) &= 4 \log p + 3 \log(1 - p) \\ \frac{d \log L(p)}{dp} &= \frac{4}{p} - \frac{3}{1 - p} \\ 0 &= -4 + 4p + 3p \\ p &= \frac{4}{7} \end{aligned}$$

In order to check if it's a maxima or a minima at $p = \frac{4}{7}$, let's check the sign on substituting p in $L''(p)$

$$\begin{aligned} \frac{d \log L(p)}{dp} &= p^3(1 - p)^2(1 - 7p) \\ \frac{d^2 \log L(p)}{dp} &= -\frac{4}{p^2} + \frac{3}{(1 - p)^2} \end{aligned}$$

Substituting $p = \frac{4}{7}$ we get a negative number, therefore, the function has a maxima at this point.

Intuitively, the maximum probability of an event occurring is equivalent to the number of times the desired result came up divided by the total number of events.

Solution 1e.

$$\begin{aligned}
f(w) &= \sum_{i=1}^n \sum_{j=1}^n (a_i^\top w - b_j^\top w)^2 + \lambda \|w\|_2^2 \\
&= \sum_{i=1}^n \sum_{j=1}^n (a_i^\top - b_j^\top)^2 w^2 + \lambda \|w\|_2^2 \quad (\dots \text{using distributive property}) \\
&= \|w\|_2^2 \sum_{i=1}^n \sum_{j=1}^n (a_i^\top - b_j^\top)^2 + \lambda \|w\|_2^2 \quad (\dots \text{using } v * v = |v|^2) \\
&= \|w\|_2^2 \left(\sum_{i=1}^n \sum_{j=1}^n (a_i^\top - b_j^\top)^2 + \lambda \right)
\end{aligned}$$

Let $\alpha = \sum_{i=1}^n \sum_{j=1}^n (a_i^\top - b_j^\top)^2$, therefore

$$\begin{aligned}
f(w) &= \|w\|_2^2 (\alpha + \lambda) \\
&= \left(\sum_{k=1}^d w_k^2 \right) (\alpha + \lambda)
\end{aligned}$$

Thus, the gradient $\nabla f(w)$ can be expressed as

$$\begin{aligned}
\nabla f(w) &= \left(\frac{\partial f(w)}{\partial w_1}, \dots, \frac{\partial f(w)}{\partial w_d} \right)^\top \\
&= (2w_1(\alpha + \lambda), \dots, 2w_d(\alpha + \lambda)) \quad \text{where } \alpha = \sum_{i=1}^n \sum_{j=1}^n (a_i^\top - b_j^\top)^2
\end{aligned}$$