

# Tarea 0

Chelsea Durazo

February 25, 2023

## 1 Optimización y probabilidad

### 1.1

Sean  $x_1, x_2, \dots, x_n$  reales que representan posiciones en una recta. Sean  $w_1, w_2, \dots, w_n$  reales positivos que representan la importancia de cada una de estas posiciones. Considera la función cuadrática

$$f(\theta) = \sum_{i=1}^n w_i(\theta - x_i)^2.$$

donde  $\theta$  es un escalar. ¿Qué valor de  $\theta$  minimiza  $f(\theta)$ ? Muestra que el óptimo que encuentres es un mínimo. ¿Qué problemas pueden surgir si algunas de las  $w_i$  son negativas?

Respuesta.

Para encontrar el mínimo de una función, primero derivamos y encontramos las raíces de la derivada para conseguir los puntos críticos:

$$f'(\theta) = \sum_{i=1}^n 2w_i(\theta - x_i).$$
$$\sum_{i=1}^n 2w_i(\theta - x_i) = 0 \quad \rightarrow \quad \theta = \frac{\sum_{i=1}^n 2w_i x_i}{\sum_{i=1}^n 2w_i}$$

Para saber si  $\theta$  es un punto mínimo, obtenemos la segunda derivada.

$$f''(\theta) = \sum_{i=1}^n 2w_i.$$

Como la segunda derivada es creciente para  $w_i > 0$  estamos hablando de un mínimo, si existen  $w_i < 0$  la segunda derivada correría el riesgo de ser negativa, y en este caso estaríamos hablando de un máximo.

### 1.2

Considera las siguientes igualdades:

$$x = (x_1, \dots, x_d) \in \mathbb{R}^d$$
$$f(x) = \max_{s \in [-1, 1]} \sum_{i=1}^d s \cdot x_i$$
$$g(x) = \sum_{i=1}^d \max_{s \in [-1, 1]} s_i \cdot x_i$$

¿Cuál de  $f(x) \leq g(x)$ ,  $f(x) = g(x)$  o  $f(x) \geq g(x)$  es verdadera para toda  $x$ ? Demuéstralo.

Respuesta:  
De  $g(x)$  tenemos:

$$s \cdot x \in \mathbb{R}^d : \quad \forall [s_i \cdot x_i], \quad [s_i \cdot x_i] \geq 0$$

Mientras que para  $f(x)$ :

$$s \in \mathbb{R} \cdot x \in \mathbb{R}^d : \quad \exists [s \cdot x_i], \quad [s \cdot x_i] \leq 0$$

Por lo tanto, al sumar todos los elementos de  $s \cdot x$ ,  $f(x) \geq g(x)$

### 1.3

Supongamos que lanzas repetidamente un dado honesto de seis caras hasta obtener un 1 o un 2 (e inmediatamente después detenerte). Cada vez que obtienes un 3 pierdes  $a$  puntos, cada vez que obtienes un 6 ganas  $b$  puntos. No ganas ni pierdes puntos si obtienes un 4 o 5. ¿Cuál es la cantidad esperada de puntos (como función de  $a$  y  $b$ ) que tienes al final?

Respuesta.  
Definiendo la recurrencia:

$$V(1) = 0$$

$$V(2) = 0$$

$$V(3) = -a + V(d)$$

$$V(4) = 0 + V(d)$$

$$V(5) = 0 + V(d)$$

$$V(6) = b + V(d)$$

Obtenemos el valor esperado sumando cada resultado posible y multiplicandolo por la probabilidad de que este ocurra:

$$\begin{aligned} E &= \frac{1}{6} \cdot 0 + \frac{1}{6} \cdot 0 + \frac{1}{6} \cdot (-a + V(d)) + \frac{1}{6} \cdot (0 + V(d)) + \frac{1}{6} \cdot (0 + V(d)) + \frac{1}{6} \cdot (b + V(d)) \\ &= \frac{-a + V(d)}{6} + \frac{2V(d)}{6} + \frac{b + V(d)}{6} \\ &= \frac{b - a + 4V(d)}{6} \end{aligned}$$

### 1.4

Supongamos que la probabilidad de que una moneda caiga en aguila es  $0 < p < 1$ , y que lanzamos esta moneda seis veces obteniendo S, A, A, A, S, A. Sabemos que la probabilidad de obtener esta secuencia es

$$L(p) = (1-p)ppp(1-p)p = p^4(1-p)^2$$

¿Cuál valor de  $p$  maximiza  $L(p)$ ? Muestra que este valor de  $p$  maximiza  $L(p)$ . ¿Cuál es una interpretación intuitiva de este valor de  $p$ ?

Respuesta.  
Reescribiendo  $L(p) = p^4 - 2p^5 + p^6$   
Calculando las derivadas:

$$L'(p) = 6p^5 - 10p^4 + 4p^3$$

$$L''(p) = 30p^4 - 40p^3 + 12p^2$$

Obteniendo puntos críticos:

$$6p^5 - 10p^4 + 4p^3 = 0 \quad \rightarrow \quad p = 0, p = \frac{2}{3}, p = 1$$

Evaluable la 2da derivada:

$$\begin{aligned} 30p^4 - 40p^3 + 12p^2 &= 0 & p &= 0, \\ 30p^4 - 40p^3 + 12p^2 &= -16/27 & p &= 2/3 \\ 30p^4 - 40p^3 + 12p^2 &= 2 & p &= 1, \end{aligned}$$

Obtenemos que  $p = 2/3$  es un máximo,  $p = 0$  es un punto silla y  $p = 1$  es un mínimo.

Intuitivamente, como la función es creciente a partir de  $p = 1$  diría que el valor que maximiza  $L(p)$  es aquel más cercano a infinito.

## 1.5

Supongamos que A y B son dos eventos tales que  $P(A|B) = P(B|A)$ . Sabemos también que  $P(A \cup B) = 1/2$  y que  $P(A \cap B) > 0$ . Muestra que  $P(A) > 1/4$ .

Respuesta.

$$P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$

Dado que  $P(A|B) = P(B|A)$

$$P(A) = P(B)$$

Luego por la propiedad de conjuntos  $A \cap B \subset A \cup B$

$$P(A \cap B) \leq P(A \cup B) = 1/2$$

Y dado que A y B no son mutuamente excluyentes

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 2 \cdot P(A) - P(A \cap B) = 1/2$$

$$P(A \cap B) = 2 \cdot P(A) - 1/2$$

$$0 < P(A \cap B) \leq 1/2 \rightarrow 0 < 2P(A) - 1/2 \leq 1/2$$

$$1/2 < 2P(A) \leq 1 \rightarrow 1/4 < P(A) \leq 1/2$$

## 1.6

Consideremos  $\mathbf{w} \in R^d$  (representado como un vector columna), constantes  $\mathbf{a}_i, \mathbf{b}_j \in R^d$  (también representados como vectores columna),  $\lambda \in R$  y un entero positivo  $n$ . Define la función

$$f(\mathbf{w}) = \left( \sum_{i=1}^n \sum_{j=1}^n (\mathbf{a}_i^\top \mathbf{w} - \mathbf{b}_j^\top \mathbf{w})^2 \right) + \frac{\lambda}{2} \|\mathbf{w}\|_2^2$$

donde el vector es  $\mathbf{w} = (w_1, \dots, w_d)^\top$  y  $\|\mathbf{w}\|_2 = \sqrt{\sum_{k=1}^d w_k^2} = \sqrt{\mathbf{w}^\top \mathbf{w}}$  es conocida como la norma  $L_2$ . Calcula el gradiente  $\nabla f(\mathbf{w})$ .

$$\nabla f(\mathbf{w}) = \left( \frac{\partial f(\mathbf{w})}{\partial w_1}, \dots, \frac{\partial f(\mathbf{w})}{\partial w_d} \right)^\top$$

Respuesta.

Comenzamos derivando la suma de forma separada:

## 2 Complejidad

### 2.1

Supongamos que tenemos una cuadrícula de  $n \times n$ , donde nos gustaría colocar cuatro rectángulos alineados a los ejes (es decir, los lados del rectángulo son paralelos a los ejes). No hay restricciones sobre la ubicación o tamaño de los rectángulos. Por ejemplo, es posible que todas las esquinas de un rectángulo sean el mismo punto (resultando en un rectángulo de tamaño cero) o que los cuatro rectángulos se traslapen entre sí.

¿Cuántas formas distintas hay de colocar los cuatro rectángulos en la cuadrícula? En general solo nos interesa la complejidad asintótica, entonces presenta tu respuesta de la forma  $O(n^c)$  u  $O(c^n)$  para algún entero  $c$ .

### 2.2

Supongamos que tenemos una cuadrícula de  $3 \times 3n$ . Comenzamos en la esquina superior izquierda (posición  $(1, 1)$ ) y nos gustaría alcanzar la esquina inferior derecha (posición  $(n, 3n)$ ) tomando pasos individuales hacia abajo o hacia la derecha. Supongamos que se nos provee con una función  $c(i, j) \in \mathbb{R}$  del costo de pasar por la posición  $(i, j)$  y que toma tiempo constante calcular cada posición. Presenta un algoritmo para calcular el costo del camino de mínimo costo desde  $(1, 1)$  hasta  $(n, 3n)$  de la manera más eficiente (con la menor complejidad en tiempo en notación  $O$  grande). ¿Cuál es el tiempo de ejecución?