Inteligencia Artificial *Juan Carlos Faz Leal*
Tarea 0: FUNDAMENTOS
27 de Enero del 2023

1. **Optimización y probabilidad.**

   a. Sean $x_1, \ldots, x_n$ reales que representan posiciones en una recta. Sean $w_1, \ldots, w_n$ 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 θ es un escalar. ¿ Qué valor de θ minimiza $f(\theta)$? Muestra que el óptimo que encuentres es un mínimo. ¿ Qué problemas pueden surgir si algunas de las w_i son negativas?

Para encontrar un valor de θ que minimiza $f(\theta)$ usaremos el criterio de la segunda derivada con respecto a θ para encontrar los puntos donde la curvatura de la función cambia. Como primer paso debemos obtener la primera derivada para encontrar sus puntos criticos para analizar el comportamiento de la función:

$$f'(\theta) = \frac{d[\sum_{i=1}^{n} w_i(\theta - x_i)^2]}{d\theta} = \sum_{i=1}^{n} 2w_i(\theta - x_i)$$

Igualamos $f'(\theta) = 0$ para encontrar su punto crítico.

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

Como originalmente queriamos encontrar el valor de θ que minimiza a la función despejamos θ en la ecuación (3). Multiplicamos w_i por θ y por x_i.

$$\sum_{i=1}^{n} w_i\theta - \sum_{i=1}^{n} w_i x_i = 0 \rightarrow \sum_{i=1}^{n} w_i\theta = \sum_{i=1}^{n} w_i x_i$$

Como θ en la ecuación no depende de la sumatoria podemos sacarla. Y pasamos la suma dividiendo.

$$\theta \sum_{i=1}^{n} w_i = \sum_{i=1}^{n} w_i x_i \rightarrow \theta = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

justifying Ahora calculamos la segunda derivada para obtener los puntos de inflexón.

$$f''(\theta) = \frac{d[\sum_{i=1}^{n} 2w_i(\theta - x_i)]}{d\theta} = \frac{d[\sum_{i=1}^{n} 2w_i\theta - \sum_{i=1}^{n} 2w_i x_i]}{d\theta} = \sum_{i=1}^{n} 2w_i$$

Mi conclusión para la primera pregunta es que como la segunda derivada de $f(\theta)$ es constante y positiva, ya que se suma n veces la misma expresión. Por lo tanto podemos decir que es convexa y tiene un único mínimo global. Si algunas de las wi son negativas puede haber un problema al obtener el valor de θ.

$$\theta = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} = \frac{(w_1 x_1)+(w_2 x_2)+...+(w_n x_n)}{(w_1+w_2+...+w_n)}$$

Si reescribimos la sumatoria nos da el valor de θ de la ecuación (7). Y podemos probar que pasa si algunos w_i son negativos.

Ejemplo. Probemos: $w_1 = 5, w_2 = -5, x_1 = 1, x_2 = -1$.

$$\theta = \frac{(5 \cdot 1)+(-5 \cdot -1)}{(5-5)}$$

Como se puede ver en la equación (8) que si algunas w_i son negativas el denominador se puede hacer 0, como consecuencia no obtendremos el punto critico.

b. Considera las siguientes igualdades.

$$x = (x_1, ..., x_d) \in \mathbb{R}^d,$$

$$f(x) = \max_{s \in [-1,1]} \sum_{i=1}^{d} s x_i,$$

$$g(x) = \sum_{i=1}^{d} \max_{s_i \in [-1,1]} s_i x_i.$$

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

Demuéstralo.

Para $f(x)$ podemos observar el valor de s que maximiza la función siempre es el mismo al multiplicarse por la sumatoria.

$$f(x) = \max_{s \in [-1,1]}(s x_1 + s x_2 + ... + s x_d) = \max_{s \in [-1,1]} s(x_1 + x_2 + ... + x_d)$$

Por otra parte en $g(x)$ el valor de s que maximiza la función cambia a cada iteración de la sumatoria.

$$g(x) = \max_{s_1 \in [-1,1]} s_1 x_1 + \max_{s_2 \in [-1,1]} s_2 x_2 + ... + \max_{s_d \in [-1,1]} s_d x_d$$

Hasta el momento podemos decir que por lo menos no se parecen y puede que no sean iguales, asi que podemos probar dando valores para determinar cual es mayor y cual es menor.

Ejemplo $f(x)$. Probemos: $s = 1, x_1 = 1, x_2 = 2$

$$f(x) = 1(1+2) = 3 \rightarrow 3 = 1(x) \rightarrow x = 3$$

Ejemplo $g(x)$. Probemos: $s_1 = 1, s_2 = 1$ y $x_1 = 1, x_2 = 2$

$$g(x) = (1)(1) + (1)(2) = 1 + 2 = 3$$

$$\rightarrow 3 = s_1 x + s_2 x \rightarrow 3 = x(s_1 + s_2) \rightarrow \frac{3}{1+1} = x \rightarrow x = \frac{3}{2}$$

justifying Con esto podemos concluir que si consideramos a x como el tiempo diremos que $g(x)$ tarda menos en llegar a la altura 3 o $y = 3$ por lo tanto $g(x)f(x)$ en este caso.

c. 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?

De la siguiente definición obtendremos la respuesta al problema. Supongamos:

$$V_n(i),$$

donde V el número de puntos obtenidos, n es el número de tiradas e i el número que salió en el dado.

Por lo tanto nuestro caso base quedaría de la siguiente manera: $V_0(i) = 0$ (No se ha lanzado el dado todavía por tanto tenemos 0 puntos). Y como queremos el valor esperado, diremos que es igual a: $E[V_0(i)] = 0$. ¿Qué pasa si lanzamos n veces el dado?

Casos: Cae 1 o 2, cae 3, cae 4 o 5 y cae 6.

$$V_n(1) = V_n(2) = 0, V_n(3) = -a + V_{n-1}, V_n(4) = V_n(5) = V_{n-1}, V_n(6) = b + V_{n-1}$$

Como dice nuestro problema: cuando cae un 1 o un 2 el juego termina por lo tanto devolvemos 0, cuando cae un 3 pierdes a puntos y sigues jugando, el jugar lo representamos como V_n entonces la siguiente jugada es V_{n-1}, cuando cae un 4 o 5 no ganas, ni pierdes y sigues jugando y cuando cae un 6 ganas b puntos y sigues jugando. Con estos casos de prueba obtendremos nuestro valor esperado de V_n, el cual se escribe como:

$$E[V_n] = \sum_{i=1}^{6} P(i) \cdot E[V_{n-1}(i)]$$

Donde el valor esperado de V_n es el producto de una sumatoria de probabilidad de que salga un número en el dado $P(i)$ con $i \in [1, 6]$ por la esperanza del siguiente lanzamiento $E[V_{n-1}]$. Sustituimos teniendo en cuenta que la probabilidad de que salga cualquier número sea $\frac{1}{6}$ y los $V_n n - 1$ de las ecuaciónes (18).

$$E[V_n] = \frac{0}{6} + \frac{0}{6} + \frac{-a+E[V_{n-1}]}{6} + \frac{E[V_{n-1}]}{6} + \frac{E[V_{n-1}]}{6} + \frac{b+E[V_{n-1}]}{6}$$

$$E[V_n] = \frac{-a+E[V_{n-1}]}{6} + \frac{2E[V_{n-1}]}{6} + \frac{b+E[V_{n-1}]}{6}$$

$$E[V_n] = \frac{b-a}{6} + \frac{2E[V_{n-1}]}{3} \to 3E[V_n] = \frac{3(b-a)}{6} + 2E[V_{n-1}]$$

$$3E[V_n] - 2E[V_{n-1}] = \frac{b-a}{2} \to E[V_n] = \frac{b-a}{2}$$

justifying Entonces, el número esperado de puntos cuando termina el juego es $\frac{b-a}{2}$

d. Supongamos que la probabilidad de que una moneda caiga en águila 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?

Derivamos y factorizamos para igualar a 0.

$$L'(p) = 4p^3(1-p)^2 - 2p^4(1-p) = 2p^3(3p-2)(p-1)$$

$$2p^3(3p-2)(p-1) = 0$$

Entonces los puntos criticos de L son: $p = 0, p = \frac{2}{3}$ y $p = 1$. Obtenemos la segunda derivada y evaluamos los puntos criticos.

$$L''(p) = \frac{d(6p^5 - 10p^4 + 4p^3)}{dp} = 30p^4 - 40p^3 + 12p^2$$

$$L''(0) = 30(0)^4 - 40(0)^3 + 12(0)^2 = 0$$

$$L''(2/3) = 30(\tfrac{2}{3})^4 - 40(\tfrac{2}{3})^3 + 12(\tfrac{2}{3})^2 = \frac{-16}{27}$$

$$L''(1) = 30(1)^4 - 40(1)^3 + 12(1)^2 = 2$$

Por tanto el valor de p que maximiza $L(p)$ es $p = \frac{2}{3} = \frac{4}{6}$. La interpretación que se le puede dar al valor de p es que de las 6 veces que se lanzó la moneda cuatro fueron aguila y dos sol.

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

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

Si sustituimos el teorema de bayes en la igualdad del ejercicio, por probabilidad condicional si $P(A|B)$ es igual a $P(B|A)$ entonces $P(A)$ y $P(B)$ tienen la misma probabilidad.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Para este caso utilizaremos la formula de eventos excluyentes y reemplazaremos el valor de $P(A \cup B)$ que es igual a $\frac{1}{2}$. También mencioné anteriormente $P(A) = P(B)$, entonces cambiamos $P(B)$.

$$\tfrac{1}{2} = P(A) + P(A) - P(A \cap B) \rightarrow P(A) = \tfrac{\frac{1}{2} + P(A \cap B)}{2}$$

Probemos para $P(A \cap B) = 1$.

$$P(A) = \tfrac{\frac{1}{2}+1}{2} = \tfrac{3}{4}$$

Por lo tanto $P(A) > \frac{1}{4}$.

f. Consideremos $w \in R^d$ (representado como un vector columna), constantes $a_i, b_j \in R^d$ (también representados como vectores columna), $\lambda \in R$ y un entero positivo n. Define la función

$$f(w) = (\textstyle\sum_{i=1}^{n} \sum_{j=1}^{n} (a_i^\top w - b_j^\top w)^2) + \frac{\lambda}{2} \|w\|_2^2,$$

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

$$f'(w) = \tfrac{d(\sum_{i=1}^{n} \sum_{j=1}^{n} (a_i^\top w - b_j^\top w)^2)}{dw} + \tfrac{d(\frac{\lambda}{2} \|w\|_2^2)}{dw}$$

$$f'(w) = \textstyle\sum_{i=1}^{n} \sum_{j=1}^{n} 2(a_i^\top w - b_j^\top w)(a_i^\top - b_j^\top) + \tfrac{d(\frac{\lambda}{2}(\sqrt{\sum_{k=1}^{d} w^2})^2)}{dw}$$

$$f'(w) = \textstyle\sum_{i=1}^{n} \sum_{j=1}^{n} 2(a_i^\top w - b_j^\top w)(a_i^\top - b_j^\top) + \tfrac{d(\frac{\lambda}{2} \sum_{k=1}^{d} w_k^2)}{dw}$$

$$\nabla f(w) = \textstyle\sum_{i=1}^{n} \sum_{j=1}^{n} 2(a_i^\top w - b_j^\top w)(a_i^\top - b_j^\top) + \lambda \sum_{k=1}^{d} w_k$$

2. **Complejidad.** Al diseñar algoritmos, es útil poder hacer cálculos rápidos y detallados para ver cuánto tiempo o espacio necesita.

   a. Supongamos que tenemos una cuadricula 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 rectangulo sean el mismo punto (resultando en un rectangulo de tamaño cero) o que los cuatro rectángulos se traslapen entre si. ¿Cuántas formas distintas hay de colocar los cuatro rectangulos en la cuadricula?

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.

Simplifiquemos primero para un rectàngulo, este se puede colocar donde sea ya que no hay restricciones, por lo que puede moverse $n \times n$ entonces su complejidad es $O(n^2)$, y también para los 3 restantes.

Entonces en total quedaría lo siguiente:

$$\text{Complejidad total} = O(n^2)^4 = O(n^8)$$

b. 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 $c(i,j) \in R$ 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?

   Para calcular nuestra función de costo utilizaremos un algoritmo de la ruta mínima el cual utiliza recurrencia, con la que se puede calcular el costo mínimo de llegar a cada posición de manera incremental. En la cuadrícula solo podemos movernos hacia abajo o a la derecha, por lo que costo actual seria:

$$cc[i][j] = c(i,j) + min(cc[i-1][j], cc[i][j-1])$$

   el minimo de llegar a cualquiera de las dos posiciones vecinas más el costo de pasar por la posición (i,j). Con el minimo nos aseguramos de estar siempre seleccionando el camino de menor coste para llegar a la posición actual. El algoritmo utiliza una matriz de tamaño $n \times 3n$ para almacenar los costos mínimos de llegar a cada posición (i,j) a partir de la posición inicial $(1,1)$.

   Entonces nuestra complejidad es $O(n \cdot 3n) = O(3n^2)$.