Inteligencia artificial
Lic. en Ciencias de la Computación
Tarea 0

Expediente: 220219356
Nombre: Juan Daniel Garcia Ruiz
Colaboradores: N/A

Al entregar esta tarea, declaro que todas las respuestas son producto de mi propio trabajo y de las personas que colaboraron especificadas arriba.

Optimización y probabilidad

1. Sean $x_1, \dots, x_n$ números reales representando posiciones sobre una recta. Sean $w_1, \dots, w_n$ números reales positivos representando la *importancia* de cada una de estas posiciones. Considera la función cuadrática,
$$f(\theta) = \sum_{i=1}^{n} w_i \left(\theta - x_i\right)^2$$
y que θ es un escalar. ¿Qué valor de θ minimiza $f(\theta)$? Muestra que el óptimo que encontraste es realmente un mínimo. ¿Qué cuestiones problemáticas pueden surgir si algunas de las w_i son negativas?

Solución: Si tenemos una función derivable $f(\theta)$, el valor de θ de los puntos de inflexión (mínimos y máximos) son las soluciones de la ecuación $f'(\theta) = 0$ Donde $f'(\theta)$ Es la derivada de f:

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

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

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

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

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

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

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

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

$$2\sum_{i=1}^{n}[w_i\theta - w_ix_i] = 0$$

$$2\sum_{i=1}^{n} w_i\theta - 2\sum_{i=1}^{n} w_ix_i$$

$$2\sum_{i=1}^{n} w_i\theta = 2\sum_{i=1}^{n} w_ix_i$$

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

El valor de θ minimiza la funcion $f(\theta)$ es la media de las posiciones x_i con pesos w_i

Si obtenemos la segunda derivada podemos observar que es positivo el cual nos indica que efectivamente es un minimo.

2. Considera las siguientes funciones,

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

$$g(\boldsymbol{x}) = \sum_{i=1}^{d} \min_{s_i\in[-1,1]} s_ix_i$$

donde $\boldsymbol{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ es un vector real y $[-1, 1]$ el intervalo cerrado entre -1 a 1. ¿Cuál de las siguientes desigualdades es cierta para toda $\boldsymbol{x}$? Demuéstralo.

$$f(\boldsymbol{x}) \leq g(\boldsymbol{x})$$
$$f(\boldsymbol{x}) = g(\boldsymbol{x})$$
$$f(\boldsymbol{x}) \geq g(\boldsymbol{x})$$

Solución: Para averiguar cual desigualdad es correcta, debemos analizar ambas funciones. En el caso de $f(x)$ sumamos todos los terminos primero y luego tomamos el minimo de la suma. Mientras que en la funcion $g(x)$ tomamos el minimo antes de sumar.

Utilizando un ejemplo el cual usamos los mismos valores pero cambiamos los signos a $2, 5$ para comparar con $d = 2$

$$
\begin{aligned}
2,5: f(\boldsymbol{x}) &= \min_{s\in[-1,1]} s(2+5) = -7 \\
g(\boldsymbol{x}) &= \min_{s_i\in[-1,1]} (s_1 * 2) + \min_{s_i\in[-1,1]} (s_2 * 5) = -7 \\
-2,-5: f(\boldsymbol{x}) &= \min_{s\in[-1,1]} s((-2)+(-5)) = -7 \\
g(\boldsymbol{x}) &= \min_{s_i\in[-1,1]} (s_1(-2)) + \min_{s_i\in[-1,1]} (s_2(-5)) = -7 \\
-2,5: f(\boldsymbol{x}) &= \min_{s\in[-1,1]} s((-2)+5) = -3 \\
g(\boldsymbol{x}) &= \min_{s_i\in[-1,1]} (s_1(-2)) + \min_{s_i\in[-1,1]} (s_2 * 5) = -7
\end{aligned}
$$

Como podemos notar:

$$f(\boldsymbol{x}) \geq g(\boldsymbol{x})$$

Esto es porque al minimizar cada termino y luego sumar com $g(x)$, no puede dar un resultado mayor a comparacion de sumar todos los terminos y luego tomar el minimo. $g(x)$ Es el minimo mas posible a comparacion con$f(x)$.

3. Supongamos que lanzas repetidamente un dado justo de seis caras hasta que obtienes un resultado de 1 (y luego te detienes). Cada vez que lanzas un 3 ganas a puntos, y cada vez que lanzas un 6 pierdes b puntos. No ganas ni pierdes puntos si lanzas un 2, 4 o 5. ¿Cuál es la cantidad de puntos (como función de a y b) que esperamos tener cuando te detengas?

Solución: Entendemos que el juego del dado de seis caras termina una vez que rodamos un 1. Y al final sumamos los puntos si rodamos un 3 y restamos los puntos si lanzamos un 6 y no haay cambios al lanzar un $2, 4, 5$.

Utilizaremos V como el valor total de la cantidad esperado de puntos. Los cuales no son $1, 3, 6$ hace que el juego continue. Si se lanza un 3 se ganan a puntos, si se lanza un b se pierden b puntos. Tenemos las siguiente funcion de V:

$$V = \frac{1}{6}0 + \frac{1}{6}(V+a) + \frac{3}{6}V + \frac{1}{6}(V-b)$$

En el cual hay una probabilidad de $\frac{1}{6}$ en obtener un 1. Una probabilidad de $\frac{1}{6}$ de obtener un 3 y sumar a puntos. Una probabilidad de $\frac{3}{6}$ de obtener un $2, 4, 5$ el cual no suma puntos. Por ultimo $\frac{1}{6}$ de botener un 6 que restaria b puntos.

Simplificando la ecuacion V nos queda como:

$$V = a - b$$

Siendo V el valor total de la cantidad esperada de puntos. Esto indica que el promedio mayoria de las veces seria ganar a puntos y perder b puntos ya que cualquier otro numero fuera de 3 o 6 no afectara el total de nuestro valor.

4. Supongamos que la probabilidad de que una moneda caiga en águila es p (donde $0 < p < 1$), y que lanzas esta moneda cinco veces obteniendo (S, A, A, A, A). Sabemos que la probabilidad de obtener esta secuencia es,

$$L(p) = (1 - p)pppp = p^4(1 - p)$$

¿Qué 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?

Solución: Para encontrar el valor p que maximiza L(P) tomaremos la derivada de P para encontrar donde habra un maximo o minimo:

$$L(p) = (1 - p)pppp = p^4(1 - p)$$
$$L(p) = p^4(1 - p)$$
$$L'(p) = \frac{d}{dp}[(1 - p)p^4] = 4p^3 - 5p^4$$

Igualando a cero nos queda:

$$4p^3 - 5p^4 = 0$$
$$p^3(4 - 5p) = 0$$

Los dos puntos criticos que obtenemos son $p = 0$ y $p = \frac{4}{5}$. Utilizaremos $p = \frac{4}{5}$ ya que en el contexto del problema, $p = 0$ no tiene sentido utilizarlo.

Para demostrar que es un maximo como es una funcion de probabilidad que aumenta hasta el punto $p = \frac{4}{5}$ y disminuye y tambien es el unico punto critico podemos decir que $p = \frac{4}{5}$ es el maximo.

Si en la secuencia de cinco lanzamientos obtenemos 4 aguilas y un solo sello pues es de esperarse que obtengamos aguila $\frac{4}{5}$ ves lo cual es igual a el valor que tenemos para p.

5. Supongamos que A y B son dos eventos tales que $P(A \mid B) = P(B \mid A)$. También sabemos que $P(A \cup B) = \frac{1}{3}$ y que $P(A \cap B) > 0$. Muestra que $P(A) > \frac{1}{6}$.

Solución: Para mostrar que $P(A) > \frac{1}{6}$ utilizaremos la definicion de probabilidad condicional y las propiedades de probabilidad.

Tenemos la probabilidad condicional:

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

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

$P(A \mid B) = P(B \mid A)$

$P(B) = P(A)$

Utilizando la formula de la probabilidad de la union de los dos eventos obtenemos:

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ $P(A \cup B) = \frac{1}{3}$ $P(A \cap B) > 0$ $\frac{1}{3} = P(A) + P(A) - P(A \cap B)$ $\frac{1}{3} = 2P(A) - P(A \cap B)$

Como $P(A \cap B) > 0$ Tenemos que $2P(A) > \frac{1}{3}$ Por lo tanto $P(A) > \frac{1}{6}$

6. Considera un vector columna $\boldsymbol{w} \in \mathbb{R}^d$ y vectores columna constantes $\boldsymbol{a}_i, \boldsymbol{b}_j \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ y un entero positivo n. Define la función con valor escalar,

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

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

Solución: Para encontrar la gradiente de la funcion calculamos la gradiente de la primera parte, es decir:

$$(\boldsymbol{a}_i^\top \boldsymbol{w} - \boldsymbol{b}_j^\top \boldsymbol{w})^2$$

Obtenemos con respecto a w:

$$2(\boldsymbol{a}_i^\top \boldsymbol{w} - \boldsymbol{b}_j^\top \boldsymbol{w})(\boldsymbol{a_i} - \boldsymbol{b_j})$$

Luego obtenemos la gradiente de la segunda parte:

$$\frac{\lambda}{2} \left\| \boldsymbol{w} \right\|_2^2$$

Nos queda con respecto a w:

$$\lambda \boldsymbol{w}$$

El cual nos da:

$$\nabla f(\boldsymbol{w}) = \sum_{i=1}^{n} \sum_{j=1}^{n} 2(\boldsymbol{a}_i^\top \boldsymbol{w} - \boldsymbol{b}_j^\top \boldsymbol{w})(\boldsymbol{a_i} - \boldsymbol{b_j}) + \lambda \boldsymbol{w}$$

Complejidad computacional

1. Supongamos que tienes una cuadrícula de puntos de $n \times n$, donde nos gustaría colocar 3 rectángulos alineados a los ejes (los lados del rectángulo son paralelos a los ejes). Cada esquina de cada rectángulo debe ser uno de los puntos en la cuadrícula, pero fuera de eso no hay restricciones sobre la ubicación o tamaño de los rectángulos. Por ejemplo, es posible que las

cuatro esquinas de un rectángulo estén en el mismo punto (resultando en un rectángulo de tamaño 0), o que todos los 3 rectángulos estén encimados. ¿De cuántas maneras se pueden colocar los 3 rectángulos sobre la cuadrícula? En general, solo nos importa la complejidad asintótica, entonces escribe tu respuesta de la forma $O(n^c)$ o de la forma $O(c^n)$ para algún entero c.

Solución:

2. Supongamos que tienes una cuadrícula de puntos de $n \times 2n$. Comenzamos en el punto de la esquina superior izquierda (el punto en la posición $(1, 1)$), y nos gustaría llegar al punto de la esquina inferior derecha (el punto en la posición $(n, 2n)$) moviéndose exclusivamente o hacia abajo o hacia la derecha. Supongamos que se nos provee una función $c(i, j)$ que produce el costo asociado con la posición (i, j), y supongamos que para cada posición toma tiempo constante calcular este costo. El costo puede ser negativo. Define el costo de un camino como la suma de $c(i, j)$ para todos los puntos (i, j) sobre el camino, incluyendo ambos extremos. Presenta un algoritmo para calcular el costo del camino de costo mínimo desde $(1, 1)$ hasta $(n, 2n)$ de la manera más eficiente posible (con la complejidad en tiempo más pequeña). ¿Cuál es el tiempo de ejecución?

Solución:

Consideraciones éticas

1. Una empresa de inversión desarrolla un modelo simple de aprendizaje automático para predecir si es probable que un individuo incumpla con un préstamo a partir de una variedad de factores, incluida la ubicación, la edad, la puntuación crediticia y los registros públicos. Después de examinar sus resultados, se encuentra que el modelo predice principalmente en función de la ubicación y que el modelo acepta principalmente préstamos de centros urbanos y niega préstamos a solicitantes rurales. Además, al observar el género y el origen étnico de los solicitantes, se encuentra que el modelo tiene una tasa de falsos positivos significativamente mayor para los solicitantes negros y masculinos que para otros grupos. En una predicción falsa positiva, un modelo clasifica erróneamente a alguien que no incumple como probable que incumpla.

Solución: El modelo puede ser que tome prejuicios en contra de grupos protegidos y tambien violando principios de igualdad y de no discriminacion. Tambien la empresa necesita ser transparente como es que el model genera las predicciones y tambien debe de ser responsable de los impactos negativos que sucedan.

2. La estilometría es una forma de predecir la autoría de un texto anónimo o impugnado, mediante el análisis de los patrones de escritura en el texto anónimo y otros textos escritos por los autores potenciales. Recientemente, se han desarrollado algoritmos de aprendizaje automático de alta precisión para esta tarea. Si bien estos modelos se utilizan normalmente para analizar documentos históricos y literatura, podrían usarse para desanonimizar una amplia gama de textos, incluido el código.

Solución: El modelo de aprendizaje tiene la capacidad de ser utilizado para desanomanizar textos lo cual seria una violacion de privacidad y de seguridad de los autores que quieren permanecer anonimos. Tambien el tipo de consentimiento contra los autores. La empresa debe de asegurar que se tenga todo el consentimiento de los autores de el cual sus escrituras se utilizan para entrenar al modelo.

3. Un grupo de investigación analizó millones de rostros de celebridades de las imágenes de Google para desarrollar una tecnología de reconocimiento facial. Las celebridades no dieron permiso para que sus imágenes se utilizaran en el conjunto de datos y muchas de las imágenes tienen derechos de autor. Para fotografías con derechos de autor, el conjunto de datos proporciona enlaces URL a la imagen original junto con cuadros delimitadores para la cara.

Solución: Para empezar utilizar el reconocimiento facial de rostros de celebridades sin su consentimiento esta mal. Esto seria una violacion de privacidad y de derechos indivivuales. La empresa debe de asegurarse en tener el consentimiento para utilizar esas imagenes y como van a ser utilizados. Tambien se debe de proteger los datos para prevenir mal uso de aquellas.

4. Los investigadores han creado recientemente un modelo de aprendizaje automático que puede predecir especies de plantas automáticamente y directamente a partir de una sola fotografía. El modelo fue entrenado usando fotografías cargadas en una aplicación por usuarios que dieron su consentimiento para usar sus fotografías con fines de investigación, y el modelo solo se usa dentro de la aplicación para ayudar a los usuarios a identificar plantas que podrían encontrar en la naturaleza.

Solución: En el caso que las potografias fueron utilizadas con el consentimiento de los usuarios y el modelo se utiliza con fines educativos hace pensar que realmente no hay mucha preocupacion etica. Lo unico que si pudiera ocasionar seria gracias a estos nuevos datos para encontrar nuevas especies seria la perturbacion de los habitats. Tambien se debe de hacer transparencia de que uso y el proposito es de estos datos.

Programación

Solución: Incorporada en `tarea0.py`.