Tarea 1. Optimización

Oscar Esaú Peralta Rosales

2/11/2020

1. Sea $f_1(x_1, x_2) = x_1^2 - x_2^2$, $f_2(x_1, x_2) = 2x_1x_2$. Representa los conjuntos de nivel asociados con $f_1(x_1, x_2) = 12$ y $f_2(x_1, x_2) = 16$ en la misma gráfica usando python. Indica sobre la gráfica, los puntos $x = [x_1, x_2]^T$ para los cuales $f(x) = [f_1(x_1, x_2), f_2(x_1, x_2)]^T = [12, 16]^T$.

Solución:

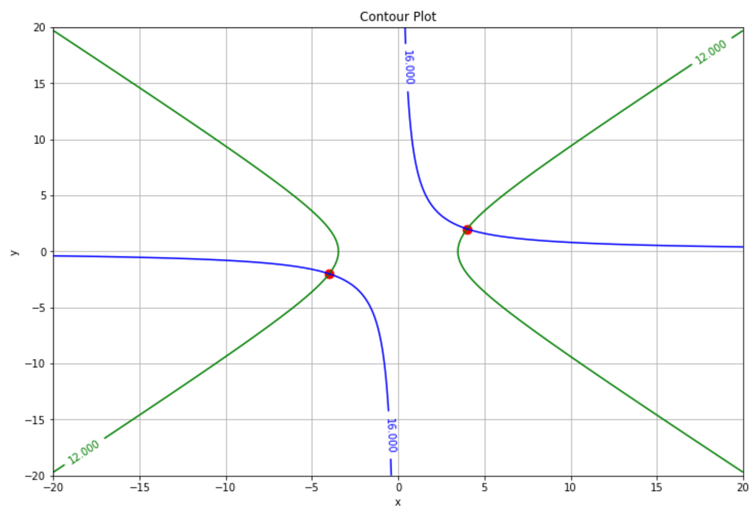

Figure 1: Contornos de nivel para las funciones f1 y f2

Nota: El archivo fuente se encuentra en el zip adjuunto.

2. Considera la función $f(x) = (a^T x)(b^T x)$, dónde a, b y x son vectores n-dimensionales. Calcula el gradiente $\nabla f(x)$ y el Hessiano $\nabla^2 f(x)$.

Solución:

$$f(x) = (a^T x)(b^T x)$$

Calculamos la primera derivada mediante la regla de la cadena

$$Df(x) = (a^T x)D(b^T x) + (b^T x)D(a^T x)$$
$$Df(x) = a^T x b^T + b^T x a^T$$

El gradiente está dado por $\nabla f(x) = (Df(x))^T$, así

$$\nabla f(x) = (a^T x b^T + b^T x a^T)^T$$
$$\nabla f(x) = a x^T b + b x^T a$$

El hessiano esta dado por $\nabla^2 f(x)$, calculamos primero la segunda derivada de $f(x)$

$$D^2 f(x) = D(a^T x b^T + b^T x a^T)$$

Puesto que $a^T x$ y $b^T x$ son productos puntos podemos rescribir lo como

$$D^2 f(x) = D(x^T a b^T + x^T b a^T)$$

Factorizando x^T tenemos

$$D^2 f(x) = D(x^T (a b^T + b a^T))$$

$$D^2 f(x) = a b^T + b a^T$$

Así el hessiano es

$$\nabla^2 f(x) = (a b^T + b a^T)^T = a b^T + b a^T$$

3. Sea $f(x) = \frac{1}{1+e^{-x}}$ y $g(z) = f(a^T z + b)$ con $||a||_2 = 1$. Muestra que $D_a g(z) = g(z)(1 - g(z))$.

Solución:

Calculamos la derivada direccional de $D_a g(z)$

$$D_a g(z) = Df(a^T z + b) D(a^T z + b) a$$
$$D_a g(z) = Df(a^T z + b)(D(a^T z) + Db)) a$$
$$D_a g(z) = Df(a^T z + b) a^T a = Df(a^T z + b)$$

la derivada de $f(x) = \frac{1}{1+e^{-x}}$ es $f'(x) = \frac{e^{-x}}{(1+e^{-x})^2}$, sea $w = a^T z + b$, luego

$$D_a g(z) = Df(w) = \frac{e^{-w}}{(1+e^{-w})^2}$$
$$D_a g(z) = \frac{1}{(1+e^{-w})} . \frac{e^{-w}}{(1+e^{-w})} = \frac{1}{(1+e^{-w})} . \frac{1+e^{-w}-1}{(1+e^{-w})}$$
$$D_a g(z) = \frac{1}{(1+e^{-w})} . \Big(1 - \frac{1}{(1+e^{-w})}\Big)$$

Por definicion $g(z) = f(w) = f(a^T z + b)$, así

$$D_a g(z) = g(z)(1 - g(z))$$

4. Calcula el gradiente de

$$f(\theta) = \frac{1}{2}\sum_{i=1}^{n}\Big[g(x_i) - g(Ax_i + b)\Big]^2$$

con respecto de θ, donde $\theta = [a_{11}, a_{12}, a_{21}, a_{22}, b_1, a_2]$, $x \in \mathcal{R}^2$, $A \in \mathcal{R}^{2x2}$, $b \in \mathcal{R}^2$ son definidos como sigue:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

$$b = \begin{pmatrix} b_1 & b_2 \end{pmatrix}^T$$

y $g : \mathcal{R}^2 \to \mathcal{R} \in \mathcal{C}^1$.

Solución

Calculamos la derivada $Df(\theta)$ mediante la regla de la cadena

$$Df(\theta) = \frac{1}{2}\sum_{i=1}^{n} 2\big[g(x_i) - g(Ax_i + b)\big]\big[Dg(x_i) - Dg(Ax_i + b)D[Ax_i + b]\big]$$

Calculamos la derivada para $D[Ax_i + b]$ con respecto a θ

$$Ax_i + b = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} x_{i1} \\ x_{i2} \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_{11}x_{i1} + a_{12}x_{i2} + b_1 \\ a_{21}x_{i1} + a_{22}x_{i2} + b_2 \end{pmatrix}$$

$$D[Ax_i + b] = \begin{pmatrix} x_{i1} & x_{i2} & 0 & 0 & 1 & 0 \\ 0 & 0 & x_{i1} & x_{i2} & 0 & 1 \end{pmatrix}$$

llamemos W a $D[Ax_i + b]$. Sea $G = [g_1, g_2]$ la derivada de $Dg(Ax_i + b)$ al usar la regla de la cadena, entonces

$$Df(\theta) = \frac{1}{2}\sum_{i=1}^{n} 2\big[g(x_i) - g(Ax_i + b)\big]\big[Dg(x_i) - GW\big]$$

Así la gradiente queda determinada por

$$\nabla f(\theta) = \frac{1}{2}\sum_{i=1}^{n} 2\big[g(x_i) - g(Ax_i + b)\big]\big[Dg(x_i) - (GW)^T\big]$$

5. Muestra que $k(A) \geq 1$ dónde $||A|| = max_x \frac{||A||}{||x||}$. Tip: Muestra que $||AB|| \leq ||A||||B||$.

Solución

Demostremos primero que $||AB|| \leq ||A||||B||$. Supongamos un x, vector de dimensión tal que Ax es una operación válida y que $||Ax|| \leq ||A||||x||$ es falso, entonces definamos:

$$||Ax|| > ||A||||x||$$

por demostrar que es falso, luego

$$\frac{||Ax||}{||x||} > ||A||$$

pero por definición $||A|| = max_x \frac{||A||}{||x||}$ por lo que $\frac{||Ax||}{||x||}$ no puede ser mayor a $||A||$. Así $||Ax|| \leq ||A||||x||$ es verdadero.

Ahora por definición $||AB|| = max_x \frac{||ABx||}{||x||}$ luego

$$||AB|| = max_x \frac{||ABx||}{||x||} \leq max_x \frac{||A||||Bx||}{||x||} = ||A||max_x \frac{||Bx||}{||x||} \leq ||A||||B||$$

Así

$$||AB|| \leq ||A||||B||$$

Cómo $k(A) = ||A||||A^{-1}||$, entonces

$$||AA^{-1}|| \leq ||A||||A^{-1}||$$
$$||I|| \leq k(A)$$
$$1 \leq k(A)$$

Así, $k(A) \geq 1$.

6. Demuestra que $x - sin(x) = o(x^2)$ cuando $x \to 0$.

Solución:

Por definición si $f(x) = o(g(x))$ entonces $\lim_{x\to a} \frac{f(x)}{g(x)} = L$, con $L = 0$. Entonces por regla de L'Hopital

$$\lim_{x\to 0} \frac{x - sin(x)}{x^2} = \lim_{x\to 0} \frac{1 - cos(x)}{2x} = \lim_{x\to 0} \frac{sin(x)}{2} = 0$$

Así $x - sin(x) = o(x^2)$.

7. Supón que $f(x) = o(g(x))$. Muestra que para algún $\epsilon > 0$ existe un $\delta > 0$ tal que si $0 < ||x|| < \delta$, entonces $|f(x)| < \epsilon|g(x)|$, i.e., $f(x) = O(g(x))$ para un $0 < |x| < \delta$.

Solución:

Por definición si $f(x) = o(g(x))$ entonces $\lim_{x\to a} \frac{f(x)}{g(x)} = 0$.

Luego por definición del limite $\lim_{x\to a} \frac{f(x)}{g(x)} = L$ tenemos que para algún $\epsilon > 0$ existe un $\delta > 0$ tal que si $0 < |x| < \delta$ entonces

$$\left|\frac{f(x)}{g(x)} - L\right| < \epsilon$$

como $L = 0$ entonces

$$\left|\frac{f(x)}{g(x)} - 0\right| < \epsilon$$

$$\frac{|f(x)|}{|g(x)|} < \epsilon$$

$$|f(x)| < \epsilon|g(x)|$$

Así, como existe un $\epsilon > 0$ tal que $|f(x)| < \epsilon|g(x)|$ entonces por definición $f(x) = O(g(x))$.

8. Muestra que si las funciones $f : \mathcal{R}^n \to \mathcal{R}$ y $g : \mathcal{R}^n \to \mathcal{R}$ satisfacen $f(x) = -g(x) + o(g(x))$ y $g(x) > 0$ para todo $x \neq 0$, entonces para todo $x \neq 0$ suficientemente pequeño, tenemos que $f(x) < 0$.

Solución:

Tenemos que

$$f(x) = -g(x) + o(g(x))$$

ó

$$f(x) + g(x) = o(g(x))$$

Por definición de limite de little-o cuando $x \to 0$ del problema anterior, existe un $\epsilon > 0$ tal que

$$|f(x) + g(x)| < \epsilon|g(x)|$$

puesto que $g(x) > 0$ y distribuyendo el valor absoluto

$$-\epsilon g(x) < f(x) + g(x) < \epsilon g(x)$$

$$-g(x)(\epsilon + 1) < f(x) < g(x)(\epsilon - 1)$$

notemos que $-g(x)(\epsilon + 1)$ siempre es negativo puesto que $g(x), \epsilon > 0$, por tanto $g(x)(\epsilon - 1)$ debe ser menor que 0 y $\epsilon \in (0, 1)$. Así $f(x) < 0$ para un $\epsilon \in (0, 1)$.