Instituto Tecnológico de Costa Rica
Escuela de Ingeniería Electrónica
Profesor: Dr. Pablo Alvarado Moya
EL 5857 Aprendizaje Automático
I Semestre 2024

Carlos Andres Mata calderón

Entrega: **20 de febrero, 2024**

Versión: 12 de febrero de 2024

Tarea 1

* Los (num) implica la ecuación que se uso del pdf "Matrix Codebook"

II. Gradientes

Sea la matriz $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ simétrica, y sea el vector $\underline{\boldsymbol{x}} \in \mathbb{R}^n$.

1. Encuentre el gradiente $\nabla_{\underline{\boldsymbol{x}}} f(\underline{\boldsymbol{x}})$ para $f(\underline{\boldsymbol{x}}) = \frac{1}{2}\underline{\boldsymbol{x}}^\top \boldsymbol{A}\underline{\boldsymbol{x}} + \underline{\boldsymbol{b}}^\top \underline{\boldsymbol{x}}$ utilizando las propiedades de trazas y gradientes vistas en clase. (10 pts)

cómo $f(x)$ es un escalar entonce podemos tomar su traza ya que

$\operatorname{tr}\{s\} = s$; donde s es un escalar

$$= \nabla_{\underline{x}}\left(\operatorname{tr}\left\{\frac{1}{2}\underline{x}^T A \underline{x} + \underline{b}^T \underline{x}\right\}\right)$$

→ por linealidad del gradiente y la traza se llega

$$= \nabla_{\underline{x}}\left(\operatorname{tr}\left\{\frac{1}{2}\underline{x}^T A \underline{x}\right\}\right) + \nabla_{\underline{x}}\left(\operatorname{tr}\left\{\underline{b}^T \underline{x}\right\}\right)$$

→ linealidad
→ Por comutatividad

→ $\operatorname{tr}\{AB\} = \operatorname{tr}\{BA\}$ → (14)

$$= \frac{1}{2}\nabla_{\underline{x}}\left(\operatorname{tr}\left\{\underline{x}^T A \underline{x} I\right\}\right) + \nabla_{\underline{x}}\left(\operatorname{tr}\left\{\underline{x}\,\underline{b}^T\right\}\right)$$

→ $\nabla_X(\operatorname{tr}\{X^T B X C\}) = BXC + B^T X C^T$
→ (117)

→ $\nabla_X(\operatorname{tr}\{XA\}) = A^T$
→ (100)

$$= \frac{1}{2}\left(A\underline{x}\cdot \cancel{I}^{1} + A^T\underline{x}\,\cancel{I^T}^{1}\right) + (\underline{b}^T)^T$$

$\hookrightarrow (A^T)^T = A$

$\hookrightarrow$ A es simétrica

$\Rightarrow A = A^T$

$$= \frac{1}{2}\left(A\underline{x} + A\underline{x}\right) + \underline{b}$$

$$= \frac{1}{\cancel{2}}\cdot \cancel{2}\, A\underline{x} + \underline{b}$$

$$\boxed{\nabla_{\underline{x}} f(\underline{x}) = A\underline{x} + \underline{b}}$$

*Se crea un archivo llamado

gradient_part.m

que se usa para probar todo.

2. Realice una función en GNU/Octave que reciba una matriz $\boldsymbol{A}$ de tamaño 2×2 y un vector $\underline{\boldsymbol{b}}$ de dos dimensiones, y que grafique la superficie paraboloide tridimensional dada por

$$f(\underline{\boldsymbol{x}}) = \frac{1}{2}\underline{\boldsymbol{x}}^\top \boldsymbol{A}\underline{\boldsymbol{x}} + \underline{\boldsymbol{b}}^\top \underline{\boldsymbol{x}}$$

en un rango de los componentes x_1 y x_2 de $\underline{\boldsymbol{x}} = [x_1, x_2]^\top$ también dado por el usuario.

Lo más relevante de este punto es que el código esté *vectorizado*, es decir, que se evite utilizar ciclos `for`, y que todo se plantee en términos de sumas y productos de matrices y vectores (incluyendo productos de Schur cuando sea necesario).

Algunas funciones que puede utilizar son, entre otras: `sum`, `sumsq`, `vecnorm`, `dot`, `meshgrid`, `surf`, `contour`. (10 pts)

Se hace el archivo plot_paraboloid.m en la función

plot_paraboloid(A, b, range_x1, range_x2)

3. Sabiendo que el mínimo se encuentra donde el gradiente de la función es cero, encuentre dónde está el mínimo de la función. (5 pts)

como

$$\nabla_{\underline{x}} f(\underline{x}) = A\underline{x} + \underline{b}$$

$$A\underline{x} + \underline{b} = \underline{0}$$

$$A\underline{x} = -b$$

$$A^{-1}A\,\underline{x} = -A^{-1}\underline{b}$$

$$\boxed{\underline{x} = -A^{-1}\underline{b}}$$

4. Use su función para mostrar tres casos de paraboloides: (10 pts)

   4.1. Matriz $\boldsymbol{A}$ igual a la matriz identidad escalada $c\boldsymbol{I}$.

   4.2. Matriz diagonal pero con los dos elementos de la diagonal distintos.

   4.3. Matriz simétrica no diagonal, que debe ser positiva definida.

   Elija $\underline{\boldsymbol{b}}$ distinto de cero y los rangos de $\underline{\boldsymbol{x}}$, de modo que el mínimo del paraboloide sea visible.

Todo se ejecuta dentro

gradient_part.m

5. Utilice ahora la función `quiver` para mostrar además el gradiente de la función paraboloide del punto 4.3. (5 pts)

Lo mismo de arriba se utiliza

$$\nabla_{\underline{x}} f(\underline{x}) = A\underline{x} + \underline{b}$$

para calcularlo.

6. Demuestre que el gradiente de la función $f(\underline{\boldsymbol{\theta}})$ definida como (5 pts)

$$f(\underline{\boldsymbol{\theta}}) = \frac{1}{2}\|\boldsymbol{X}\underline{\boldsymbol{\theta}} - \underline{\boldsymbol{y}}\|^2$$

   respecto a $\underline{\boldsymbol{\theta}} \in \mathbb{R}^n$, está dado por

$$\nabla_{\underline{\boldsymbol{\theta}}} f(\underline{\boldsymbol{\theta}}) = \boldsymbol{X}^\top(\boldsymbol{X}\underline{\boldsymbol{\theta}} - \underline{\boldsymbol{y}})$$

   donde $\boldsymbol{X} \in \mathbb{R}^{m\times n}$. Debe utilizar las propiedades de trazas y gradientes vistas en clase.

$$\nabla_{\underline{\theta}} f(\underline{\theta}) = \nabla_{\underline{\theta}} \left(\underbrace{\frac{1}{2} \|X\underline{\theta} - \underline{y}\|^2}_{\text{f(x) es calor entonces lo transformo en una traza}} \right)$$

$$= \nabla_{\underline{\theta}} \left(\frac{1}{2} \operatorname{tr}\{ \|X\underline{\theta} - \underline{y}\|^2 \} \right)$$

$$= \frac{1}{2} \nabla_{\underline{\theta}} \left(\operatorname{tr}\{ (X\underline{\theta} - \underline{y})^T (X\underline{\theta} - \underline{y}) \} \right)$$

$\hookrightarrow \operatorname{tr}\{BA\} = \operatorname{tr}\{AB\} \Longrightarrow (14)$

$$= \frac{1}{2} \nabla_{\underline{\theta}} \left(\operatorname{tr}\{ (X\underline{\theta} - \underline{y})(X\underline{\theta} - \underline{y})^T \} \right)$$

$\hookrightarrow \nabla_X \operatorname{tr}\{ (AXB + C)(AXB + C)^T \}$

$= 2A^T (AXB + C) B^T$

$\hookrightarrow (119)$

$\Rightarrow B = I$

$$= \frac{1}{\cancel{2}} \cdot \cancel{2} \cdot X^T (X\underline{\theta} I + -\underline{y}) I^T$$

$$= \boxed{X^T (X\underline{\theta} - \underline{y})}$$

7. Encuentre ahora el gradiente $\nabla_{\boldsymbol{X}} f(\boldsymbol{X})$ con (10 pts)

$$f(\boldsymbol{X}) = \frac{1}{2}\|\boldsymbol{X}\underline{\boldsymbol{\theta}} - \underline{\boldsymbol{y}}\|^2$$

respecto a $\boldsymbol{X} \in \mathbb{R}^{m\times n}$, donde $\underline{\boldsymbol{\theta}} \in \mathbb{R}^n$. Debe utilizar las propiedades de trazas y gradientes vistas en clase.

Ayuda: *Revise propiedades adicionales de las trazas y los gradientes en el "Matrix Codebook" de Petersen y Pedersen (2012). Nótese que allí en vez de $\nabla_{\boldsymbol{X}}$, se usa $\frac{\partial}{\partial \boldsymbol{X}}$. Debe utilizar en su solución la notación indicada en la sección I, es decir $\nabla_{\boldsymbol{X}}$.*

$$= \nabla_X f(X)$$

$$= \nabla_X \left(tr\left\{ \frac{1}{2} (X\underline{\theta} - \underline{y})^T (X\underline{\theta} - \underline{y}) \right\} \right)$$

$\Rightarrow (14) \Rightarrow tr(AB) = tr(BA)$

$$= \frac{1}{2} \nabla_X \left(tr\left\{ (X\underline{\theta} - \underline{y})(X\underline{\theta} - \underline{y})^T \right\} \right)$$

$\Rightarrow \nabla_X tr\{(AXB + C)(AXB+C)^T\}$

$= 2A^T(AXB+C)B^T \Rightarrow (119)$

$\Rightarrow A = I$

$$= \frac{1}{\cancel{2}} \cdot \cancel{2}\, I^T (I X\underline{\theta} + - \underline{y})\, \underline{\theta}^T$$

$$\boxed{= (X\underline{\theta} - \underline{y})\underline{\theta}^T}$$

8. Encuentre ahora el gradiente $\nabla_{\underline{x}} f(\underline{x})$ con la composición $f(\underline{x}) = g(h(\underline{x}))$ con $g : \mathbb{R} \to \mathbb{R}$ y $h : \mathbb{R}^n \to \mathbb{R}$, ambas diferenciables. (10 pts)

g es escalar y su variable es h que es un escalar

h es escalar y su variable $\underline{x}$ un vector

tomemos la definicón para del $\nabla_{\underline{x}} f(\underline{x})$

$$\nabla_{\underline{x}} f(\underline{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

g(h) tiene una derivada corriente respecto a h

Tomemos el primer elemento

$$\frac{\partial f}{\partial x_1} = (g(h(\underline{x})))_{x_1}$$

se aplica regla de cadena

$$= g'(h(\underline{x})) \cdot \frac{\partial h(\underline{x})}{\partial x_1}$$

esto sucede para cáda elemento

$$\nabla_{\underline{x}} f(\underline{x}) = \begin{bmatrix} g'(h(\underline{x})) \cdot \frac{\partial h(\underline{x})}{\partial x_1} \\ \vdots \\ g'(h(\underline{x})) \cdot \frac{\partial h(\underline{x})}{\partial x_n} \end{bmatrix}$$

practicamente esto es decir que

$$\boxed{\nabla_{\underline{x}} f(\underline{x}) = g'(h(\underline{x})) \; \nabla_{\underline{x}} h(\underline{x})}$$

9. Sea ahora $f(\boldsymbol{x}) = g(\boldsymbol{a}^\top \boldsymbol{x})$ con $g : \mathbb{R} \to \mathbb{R}$ continuamente diferenciable y $\boldsymbol{a} \in \mathbb{R}^n$ un vector. Encuentre $\nabla_{\boldsymbol{x}} f(\boldsymbol{x})$. (5 pts)

por ser un escalar

Por lo anterior ...

$$\nabla_{\underline{x}} f(\underline{x}) = g'(\underline{a}^T \underline{x}) \; \nabla_{\underline{x}} \left(\mathrm{tr}\{ \underline{a}^T \underline{x} \} \right)$$

$\rightarrow \mathrm{tr}\{AB\} = \mathrm{tr}\{BA\}$

$\rightarrow (141)$

$$= g'(\underline{a}^T \underline{x}) \; \nabla_{\underline{x}} \left(\mathrm{tr}\{ \underline{x}\, \underline{a}^T \} \right)$$

$\rightarrow \nabla_A (\mathrm{tr}\{AB\}) = B^T$

$$= g'(\underline{a}^T \underline{x}) \cdot (\underline{a}^T)^T$$

$$\boxed{= g'(\underline{a}^T \underline{x}) \cdot a}$$

III. Matrices positivas definidas

1. Sea $\underline{z} \in \mathbb{R}^n$ un vector n-dimensional. Muestre que $\boldsymbol{A} = \underline{z}\underline{z}^\top$ es positiva semidefinida. (10 pts)

$$A = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \begin{bmatrix} z_1 & z_2 & \cdots & z_n \end{bmatrix}$$

$$A = \begin{bmatrix} z_1 z_1 & z_1 z_2 & \cdots & z_1 z_n \\ z_2 z_1 & z_2 z_2 & \cdots & z_2 z_n \\ \vdots & \vdots & \ddots & \\ z_n z_1 & z_n z_2 & & z_n z_n \end{bmatrix} = \begin{matrix} z_1^2 & z_1 z_2 \cdots & z_1 z_n \\ z_1 z_2 & z_2^2 \cdots & z_2 z_n \\ \vdots & \vdots & \ddots \\ z_{1n} & z_2 z_n & z_n^2 \end{matrix}$$

Note que la primero que A los valores

$a_{ij} = a_{ji}$ entonces A es simétrica.

Ahora revisenos que es semipositiva definida

Para que sea semi positiva definida

$\forall \underline{x} \in \mathbb{R}^n \setminus 0$ tal que

$$\underline{x}^T A \underline{x} \geq 0$$

$$\underline{x}^T (\underline{z}\, \underline{z}^T) \underline{x} \geq 0$$

$$(\underline{x}^T \underline{z}^T)(\underline{z}^T \underline{x})$$

$$\underbrace{(\underline{z}^T \underline{x})^T}\ \underbrace{(\underline{z}^T \underline{x})} \geq 0$$

Vea que ambos darán el mismo escalar
xk el producto interno dará el mismo escalar s
y cómo $s = s^T$ entonces...

$$(\underline{z}^T \underline{x})(\underline{z}^T \underline{x}) \geq 0$$

$$\boxed{(\underline{z}^T \underline{x})^2 \geq 0}$$

Esto siempre se cumple

//

2. Sea $\underline{z} \in \mathbb{R}^n$ un vector n-dimensional no nulo, y $\boldsymbol{A} = \underline{z}\underline{z}^\top$. Encuentre cuál es el espacio nulo de $\boldsymbol{A}$ y su rango. (10 pts)

el espacio nulo es

$\exists \underline{x} \in \mathbb{R}^n$

$$A\underline{x} = \underline{0}$$

$$\underline{z}\underbrace{\underline{z}^T\underline{x}}_{\text{escalar}} = \underline{0}$$

$$\underline{z}\, s = 0$$

$$\Rightarrow s = 0$$

$\boxed{\text{para } s = 0 \text{ implica } \underline{z}^T \underline{x} \text{ tienen ser ortogonales, para que cumpla que } s \text{ sea igual a } 0.}$

3. Compruebe sus hallazgos en GNU/Octave de la siguiente forma: Genere un vector $\underline{z}$ aleatorio no nulo en tres dimensiones. Genere con él la matriz $\boldsymbol{A} = \underline{z}\underline{z}^\top$. Sintetice 100 vectores aleatorios $\underline{n}_i$ en el espacio nulo. Compruebe que en cada caso $\boldsymbol{A}\underline{n}_i = \underline{0}$. (15 pts)

Sugerencia: el procedimiento de Gram-Schimdt revisado en el curso de Señales y Sistemas le puede ser de utilidad.

Se utiliza null_space.m para generar los vectores n_i

Se prueba en define_positive_matrix_part.m

IV. Ecuaciones normales

1. Usando los puntos del paraboloide en el punto 4.3, utilice las ecuaciones normales de regresión lineal para encontrar los parámetros de un plano que mejor ajuste dichos puntos. (10 pts)

2. Grafique el plano encontrado. (5 pts)

se hace todo en normal_equotion_part.m

Anexo:

Para la parte Gradiente punto 1.

No se leyó bien las instrucciones por lo que se hizo apie el equivalente de $\nabla_{\underline{x}} f(\underline{x})$. Por lo que se adjunta cómo anexo, por si gusta ver cómo se llega al mismo gradiente.

II. Gradientes

Sea la matriz $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ simétrica, y sea el vector $\underline{\boldsymbol{x}} \in \mathbb{R}^n$.

1. Encuentre el gradiente $\nabla_{\underline{\boldsymbol{x}}} f(\underline{\boldsymbol{x}})$ para $f(\underline{\boldsymbol{x}}) = \frac{1}{2}\underline{\boldsymbol{x}}^\top \boldsymbol{A}\underline{\boldsymbol{x}} + \underline{\boldsymbol{b}}^\top \underline{\boldsymbol{x}}$ utilizando las propiedades de trazas y gradientes vistas en clase. (10 pts)

$$f(\underline{x}) = \frac{1}{2}\underline{x}^T A \underline{x} + \underline{b}^T \underline{x}$$

$$= \frac{1}{2}\left(\underline{x}^T A + 2\underline{b}^T\right)\underline{x}$$

Tomar toda la primera columna de A

$$= \frac{1}{2}\left(\left[\underline{x}\cdot \underbrace{a_{:1}} \quad \underline{x}\cdot a_{:2} \ldots \underline{x}\cdot a_{:n}\right] + 2\left[b_1\ b_2 \ldots b_n\right]\right)\underline{x}$$

$$= \left(\frac{1}{2}\left[\underline{x}\cdot a_{:1} + 2b_1 \quad \underline{x}\cdot a_{:2} + b_2 \cdot 2 \ldots \underline{x}\cdot a_{:n} + b_n\cdot 2\right]\right)\underline{x}$$

$$= \frac{1}{2}\left(\underbrace{\underline{x}\cdot a_{:1}}_{1} + b_1\cdot 2)x_1 + (\underline{x}\cdot a_{:2} + b_2\cdot 2)x_2 + \cdots + (\underline{x}\cdot a_{:n} + b_n\cdot 2)x_n\right)$$

Hagomo zoom en esto para entender el producto interno de esto

① $\underline{x} \cdot a_{:1}$ $\underline{x} \cdot a_{:n}$

$= x_1 a_{11} + x_2 a_{21} + \cdots + x_n a_{n1}$ $= x_1 a_{1n} + \cdots + x_n a_{nn}$

Volviendo ...

$$= \frac{1}{2}\Big((x_1 a_{11} + x_2 a_{21} + \cdots + x_n a_{n1} + b_1 2) x_1 +$$
$$(x_1 a_{12} + x_2 a_{22} + \cdots x_n a_{n2} + b_2 2) x_2 +$$
$$\vdots \qquad +$$
$$(x_1 a_{1n} + x_2 a_{2n} + \cdots + x_n a_{nn} + b_n 2) x_n \Big)$$

$$= \frac{1}{2}\Big(x_1^2 a_{11} + x_1 x_2 a_{21} + \cdots + x_1 x_n a_{n1} + x_1 b_1 \cdot 2 +$$
$$x_2 x_1 a_{12} + x_2^2 a_{22} + \cdots + x_2 x_n a_{n2} + x_2 b_2 \cdot 2 +$$
$$\vdots \qquad +$$
$$x_n x_1 a_{1n} + x_n x_2 a_{2n} + \cdots + x_n^2 a_{nn} + x_n b_n \cdot 2 \Big)$$

Ahora tomemos la definición de $\nabla_{\underline{x}} f(\underline{x})$

$$\nabla_{\underline{x}} f(\underline{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

Apliquemos $\frac{\partial f}{\partial x_1}$

$$= \Big(\frac{1}{2} \Big(x_1^{\cancel{2}} a_{11} + \cancel{x_1} x_2 a_{21} + \dots + \cancel{x_1} x_n a_{n1} + \cancel{x_1} b_1 \cdot 2 +$$

$$x_2 \cancel{x_1} a_{12} + \cancel{x_2^2 a_{22}} + \cancel{\dots} + \cancel{x_2 x_n a_{n2}} + \cancel{x_2 b_2 \cdot 2} +$$

$$\emptyset \quad \emptyset \quad \emptyset \quad \emptyset$$

$$\vdots$$

$$+$$

$$x_n \cancel{x_1} a_{1n} + \cancel{x_n x_2 a_{2n}} + \cancel{\dots} + \cancel{x_n^2 a_{nn}} + \cancel{x_n b_n \cdot 2} \Big) \Big)_{x_1}$$

$$\emptyset \quad \emptyset \quad \emptyset \quad \emptyset$$

$$= \frac{1}{2} \Big(2 x_1 a_{11} + (x_2 a_{21} + \dots x_n a_{n1}) + b_1 \cdot 2 +$$

$$(x_2 a_{12} + x_3 a_{13} + \dots x_n a_{1n}) \Big)$$

como A es simétrica $a_{ij} = a_{ji}$ ósea
$a_{21} = a_{12}$ entonces ...

$$= \frac{1}{\cancel{2}}\left(\cancel{2}\, x_1 a_{11} + \cancel{2} x_2 a_{21} + \cancel{2} x_3 a_{31} + \dots + \cancel{2} x_n a_{n1} + b_1 \cancel{2}\right)$$

$$= x_1 a_{11} + x_2 a_{21} + x_3 a_{31} + \dots + x_n a_{n1} + b_1$$

Tomar toda la primera columna de A

$$= \underline{x} \cdot a_{:1} + b_1$$

Lo mismo para las demás entonces

$$\nabla_{\underline{x}} f(\underline{x}) = \begin{bmatrix} \underline{x}^T a_{:1} + b_1 \\ \underline{x}^T a_{:2} + b_2 \\ \vdots \\ \underline{x}^T a_{:n} + b_n \end{bmatrix}$$

$$\nabla_{\underline{x}} f(\underline{x}) = Ax + b \; ;$$

FIN DEL ANEXO