IMT2220, Cálculo para ciencia de datos, 2023-2
Tarea 1

Fecha entrega: 01 de septiembre de 2023

Instrucciones

Pueden discutir los problemas con sus compañeros. Sin embargo, sus entregas, tanto la parte escrita como los códigos, deben ser individuales. **No está permitido** copiar las respuestas de alguien más ni dejar que otros copien sus respuestas. El hacerlo se reflejará con la nota mínima en la evaluación (1.0).
Su entrega debe estar compuesta por un único archivo .pdf **escrito en LaTeX** que incluya todas las respuestas y todos los gráficos que se están pidiendo, junto con un archivo .zip que incluya todos los códigos que produjeron y utilizaron durante esta tarea. Estos dos archivos deben ser subidos en Canvas. Si hay más de una línea en un gráfico, utilice distintos estilos de línea o distintos colores para diferenciarlas e incluya leyendas. No se olvide de nombrar todos sus ejes y líneas en las leyendas. **Cada gráfico debe ser comentado** (un gráfico sin un análisis del mismo no es una respuesta apropiada).

Preguntas

1. (5 puntos) Considere $\langle \cdot, \cdot \rangle : \mathbb{R}^2 \to \mathbb{R}$ definido por

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle := x_1 y_1 - (x_1 y_2 + x_2 y_1) + 2 x_2 y_2$$

Muestre que este es un producto interno en $\mathbb{R}^2$ y muestre una fórmula para la norma inducida por él.

Solution: Probar simetría es trivial. Para la linealidad, basta expandir y agrupar términos:

$$\begin{aligned} \langle \boldsymbol{x}_1 + \alpha \boldsymbol{w}_1, \boldsymbol{y} \rangle &= (x_1 + \alpha w_1) y_1 - (x_1 + \alpha w_1) y_2 - (x_2 + \alpha w_2) y_1 + 2(x_2 + \alpha w_2) y_2 \\ &= \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \alpha \langle \boldsymbol{w}, \boldsymbol{y} \rangle \end{aligned}$$

Luego:

$$\begin{aligned} \langle \boldsymbol{x}, \boldsymbol{x} \rangle &= x_1^2 - 2x_1 x_2 + 2x_2^2 \\ &= (x_1 - x_2)^2 + x_2^2 \\ &\geq 0 \end{aligned}$$

Con esto, la norma inducida corresponde a:

$$\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle} = \sqrt{(x_1 - x_2)^2 + x_2^2}$$

2. (10 puntos) Muestre que la siguiente es una distancia en $\mathbb{R}$:

$$d(x, y) = \frac{|x - y|}{1 + |x - y|}$$

Solution: La desigualdad triangular se prueba como sigue:

$$\begin{aligned}
d(x, y) &= \frac{|x - y|}{1 + |x - y|} \\
&= 1 - \frac{1}{1 + |x - y|} \\
&= 1 - \frac{1}{1 + |x - z + z - y|} \\
&\leq 1 - \frac{1}{1 + |x - z| + |z - y|} \\
&= \frac{|x - z| + |z - y|}{1 + |x - z| + |z - y|} \\
&\leq \frac{|x - z|}{1 + |x - z|} + \frac{|z - y|}{1 + |z - y|} \\
&= d(x, z) + d(z, y)
\end{aligned}$$

El resto de las propiedades son fáciles de probar.

3. Considere el espacio de funciones continuas a valores reales en $[0, 1]$, es decir, $C([0, 1], \mathbb{R})$.

   (a) (5 puntos) Demuestre que la siguiente es una norma en este espacio:

$$\|f\| \coloneqq \max_{x \in [0,1]} |f(x)|$$

Solution: Veamos las propiedades de una norma. Sean $f, g \in C([0, 1], \mathbb{R})$ y $\lambda \in \mathbb{R}$ constante. Entonces:

1. $\|\lambda f\| = \max_{x \in [0,1]} |\lambda f(x)| = |\lambda| \, \|f\|$
2. $\|f\| \geq 0$ por definición.
3. Si $f = 0$, entonces $\max_{x \in [0,1]} |f(x)| = 0$. Si, por otro lado, $\max_{x \in [0,1]} |f(x)| = 0$, entonces $f = 0$ en $[0, 1]$.
4. $$\begin{aligned}
\|f + g\| &= \max_{x \in [0,1]} |f(x) + g(x)| \\
&\leq \max_{x \in [0,1]} (|f(x)| + |g(x)|) \\
&\leq \max_{x \in [0,1]} |f(x)| + \max_{x \in [0,1]} |g(x)| \\
&\leq \|f\| + \|g\|
\end{aligned}$$

(b) (5 puntos) Considere la sucesión de funciones $\{f_n\} \subset C([0,1],\mathbb{R})$ definidas por:

$$f_n(x) = \begin{cases} 1 - n\,x & x \in [0, 1/n] \\ 0 & x \in (1/n, 1] \end{cases}$$

Considere, por otro lado, la norma:

$$\|f\|_{L^2([0,1])} := \left(\int_0^1 |f(x)|^2 \, dx \right)^{1/2}$$

¿es cierto entonces que f_n converge a la función 0 bajo la distancia $d(f,g) = \|f-g\|_{L^2([0,1])}$?

Solution: Verifiquemos esto:

$$\begin{aligned} d(f_n, 0)^2 &= \|f_n\|^2_{L^2([0,1])} \\ &= \int_0^1 |f_n(x)|^2 \, dx \\ &= \int_0^{1/n} (1 - nx)^2 \, dx \\ &= \int_0^{1/n} (1 - 2nx + n^2x^2) \, dx \\ &= \left[x - nx^2 + \frac{1}{3} n^2 x^3 \right]_0^{1/n} \\ &= \frac{1}{n} - \frac{1}{n} + \frac{1}{3n} \\ &= \frac{1}{3n} \overset{n\to\infty}{\to} 0 \end{aligned}$$

Como la distancia al cuadrado tiende a 0, entonces la distancia tiende a 0, por lo que f_n converge a 0 bajo dicha distancia.

4. (10 puntos) Considere el espacio de matrices simétricas cuadradas de 2×2 con el producto interno de Frobenius:

$$\langle A, B \rangle_F = \sum_{i=1}^{2} \sum_{j=1}^{2} A_{ij} B_{ij}$$

Genere una base ortonormal de este espacio usando el algoritmo de Gramm-Schmidt a partir del conjunto linealmente independiente $\{A_0, A_1, A_2\}$, donde:

$$A_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad A_1 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

Solution: Las matrices resultantes debiesen ser (ortogonalizando en el orden $A_0 \to A_1 \to A_2$):

$$\hat{A}_0 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$
$$\hat{A}_1 = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix}$$
$$\hat{A}_2 = \frac{1}{\sqrt{10}} \begin{bmatrix} -2 & 1 \\ 1 & 2 \end{bmatrix}$$

5. El problema de clasificación lineal binario consiste en lo siguiente: dado un set de datos separables $\boldsymbol{x}_i \in \mathbb{R}^n, i = 1, \ldots, N$ etiquetados respectivamente por $y_i \in \{-1, 1\}, i = 1, \ldots, N$, encontrar un hiperplano que separe estos datos según sus etiquetas. Para el caso de $\mathbb{R}^2$ podemos ver por ejemplo la Figura 1, donde un set de datos etiquetados (en azul y rojo) es separado por un hiperplano (en negro).

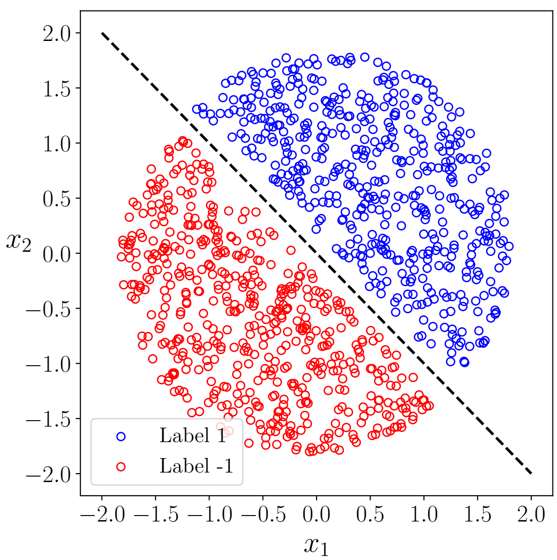

Figura 1: Ejemplo de una resolución al problema de clasificación binaria

Un algoritmo que es capaz de encontrar dicho hiperplano separador corresponde al algoritmo del perceptron.

donde $\eta > 0$ corresponde a un parámetro que representa la tasa de aprendizaje del algoritmo (y puede depende de cada iteración).

Para el resto de este problema considere que $\boldsymbol{w}_0 = 0$, $\eta = 1$, que para cada dato $\boldsymbol{x}_i$ existe $r > 0$ tal que $\|\boldsymbol{x}_i\|_2 \leq r$ para $i = 1, \ldots, N$, y que existen $\rho > 0$ y $\boldsymbol{v} \in \mathbb{R}^n$ unitario tal que:

$$\rho \leq y_i \langle \boldsymbol{v}, \boldsymbol{x}_i \rangle, \quad i = 1, \ldots, N \tag{1}$$

Algorithm 1: Perceptrón

Data: $(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_N, y_N)$, and initial guess $\boldsymbol{w}_0$
Result: $\boldsymbol{w}_{N+1} \in \mathbb{R}^n$ hiperplano separador
$\boldsymbol{w}_1 \leftarrow \boldsymbol{w}_0$
for $k = 1, \ldots, N$ **do**
  if $\operatorname{sign}(\langle \boldsymbol{x}_k, \boldsymbol{w}_k \rangle) \neq y_k$ **then**
    $\boldsymbol{w}_{k+1} \leftarrow \boldsymbol{w}_k + \eta y_k \boldsymbol{x}_k$
  else
    $\boldsymbol{w}_{k+1} \leftarrow \boldsymbol{w}_k$
  end
end

(a) (4 puntos) Denote por M el número de veces en las cuales se realiza una actualización en el algoritmo (esto está descrito por un subconjunto de los N datos). Demuestre que:

$$\rho M \leq \|\boldsymbol{w}_{N+1}\|$$

Hint: Use el hecho que si se cumple (1), entonces en particular se cumple para el conjunto de las iteraciones donde se actualiza el algoritmo.

Solution: Sea J el conjunto de pasos del algoritmo donde el algoritmo actualiza el hiperplano separador $\boldsymbol{w}$. Esto es, $|J| = M$. Así:

$$\begin{aligned}
\rho M &\leq \sum_{i \in J} y_i \langle \boldsymbol{v}, \boldsymbol{x}_i \rangle \\
&= \left\langle \boldsymbol{v}, \sum_{i \in J} y_i \boldsymbol{x}_i \right\rangle \\
&\leq \left\| \sum_{i \in J} y_i \boldsymbol{x}_i \right\| \\
&= \|\boldsymbol{w}_{N+1}\|
\end{aligned}$$

(b) (4 puntos) Demuestre que:

$$\|\boldsymbol{w}_{N+1}\| \leq \sqrt{Mr^2}$$

Hint: Use el hecho que $\|\boldsymbol{w}_{N+1}\| = \sqrt{\|\boldsymbol{w}_{N+1}\|^2}$ y escribalo como una suma telescópica.

Solution:

$$\begin{aligned}
\|\boldsymbol{w}_{N+1}\| &= \sqrt{\|\boldsymbol{w}_{N+1}\|^2} \\
&= \sqrt{\sum_{i\in J}(\|\boldsymbol{w}_{i+1}\|^2 - \|\boldsymbol{w}_i\|^2)} \\
&= \sqrt{\sum_{i\in J}(\|\boldsymbol{w}_i + y_i\boldsymbol{x}_i\|^2 - \|\boldsymbol{w}_i\|^2)} \\
&= \sqrt{\sum_{i\in J}\underbrace{2y_i\left\langle \boldsymbol{w}_i, \boldsymbol{x}_i\right\rangle}_{\leq 0} + \|\boldsymbol{x}_i\|^2} \\
&\leq \sqrt{Mr^2}
\end{aligned}$$

(c) (2 puntos) Concluya que el número de actualizaciones del algoritmo está acotado como $M \leq r^2/\rho^2$.

Solution: De (a) y (b), tenemos que

$$M\rho \leq \sqrt{M}r \Rightarrow M \leq r^2/\rho^2$$

(d) (15 puntos) Programe el algoritmo del perceptrón. Usando el set de datos `data_HW1.csv`, entrene 1000 perceptrones. Procure hacer esto cada uno con un orden distinto de los datos. Para esto es de utilidad la función de numpy `numpy.random.shuffle` para elegir un orden aleatorio. Este set de datos contiene 3 columnas:

1. las dos primeras corresponden a las coordenadas x e y de los datos,
2. la última columna corresponde a las etiquetas asignadas a cada uno de ellos.

Para cada perceptrón, calcule cuándo dejaron de actualizar el hiperplano separador y reporte estos resultados en un histograma y el máximo valor alcanzado en este histograma. ¿Qué podemos decir sobre la constante ρ si sabemos que $r = 2$? Grafique el set de datos con cinco de los hiperplanos resultantes. ¿Son todos iguales? Discuta sobre esta última pregunta.