Àlex Montoya Pérez (242873)

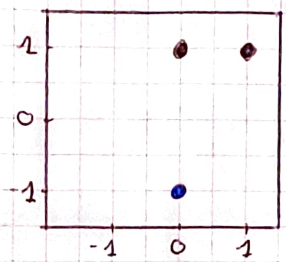
DELIVERABLE MACHINE LEARNING

LAB 7A: EX 2 (SVM)

$C_1 = \{ x_1 = \binom{0}{-1} \}$ - label 1

$C_2 = \{ x_2 = \binom{0}{1}, x_3 = \binom{1}{1} \}$ - label -1

a)



→ El nostre classificador serà el següent:

$$g\left(\binom{x_1}{x_2}\right) = w_1 \cdot x_1 + w_2 \cdot x_2 + b$$

b) Volem el mínim de $\|w\| = \sqrt{x_1^2 + \dots + x_D^2}$ subjectat a una restricció, en el nostre cas:

$$\min_w \frac{1}{2} \|w\|_2^2 \quad \text{subjectat a la restricció} \quad y^{(n)} \cdot (w^T x^{(n)} + b) \geq 1 \quad (\forall n = 1,2,3..)$$

on $\{(x^{(n)}, y^{(n)})\}$ son els parells d'entrenament.

c) $L_D(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \cdot \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \cdot x^{(i)T} \cdot x^{(j)}$

→ Per simplificar els càlculs ho podem expressar en forma de matriu tal que:

$$L_D(\alpha) = (\alpha_1 \dots \alpha_n) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} - \frac{1}{2} (\alpha_1 \dots \alpha_n) \begin{pmatrix} y^{(1)} x^{(1)T} x^{(1)} y^{(1)} & \dots & y^{(1)} x^{(1)T} x^{(N)} y^{(N)} \\ \vdots & \dots & \vdots \\ y^{(N)} x^{(N)T} x^{(1)} y^{(1)} & \dots & y^{(N)} x^{(N)T} x^{(N)} y^{(N)} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$$

en el nostre cas tindrem:

$$L_D(\alpha) = (\alpha_1\ \alpha_2\ \alpha_3) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \frac{1}{2} (\alpha_1\ \alpha_2\ \alpha_3) \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$$

→ Volem $\max_{\alpha_1 \alpha_2 \alpha_3} L_D(\alpha)$ per el que hem de $\frac{\partial L_D(\alpha)}{\partial \alpha} = \emptyset$

$$\frac{\partial L_D(\alpha)}{\partial \alpha} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \emptyset$$

obtenint el següent sistema d'equacions:

$$\left.\begin{aligned} 1 &= \alpha_1 + \alpha_2 + \alpha_3 \\ 1 &= \alpha_1 + \alpha_2 + \alpha_3 \\ 1 &= \alpha_1 + \alpha_2 + 2\alpha_3 \end{aligned}\right\}$$

} dependents, per el que afegim $\{ \sum_i y_i \alpha_i = 0 \}$

$\alpha_1 - \alpha_2 - \alpha_3 = \emptyset$

⇒ Resolem el sistema d'equacions i obtenim

$$\alpha_3 = \emptyset \,;\ \alpha_2 = \frac{1}{2} = \alpha_1$$

d) Del Primal Lagrangian podem dir:

Traduïm $\alpha \to w$ (sol. dual) (sol. primal)

$$w = \alpha_1 \cdot x_1 - \alpha_2 \cdot x_2 - \alpha_3 \cdot x_3 = \frac{1}{2}\begin{pmatrix}0\\-1\end{pmatrix} - \frac{1}{2}\begin{pmatrix}0\\1\end{pmatrix} - 0\cdot\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}0\\-1\end{pmatrix}$$

Per el que tindrem

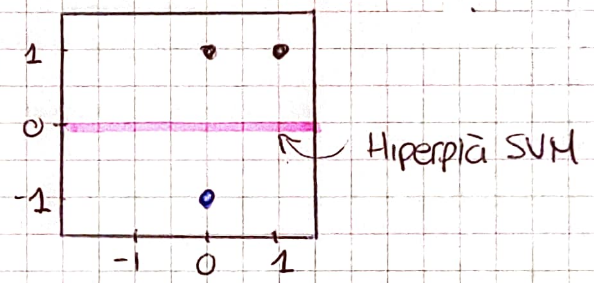
$$\left.\begin{array}{l} g(x^{(1)}) = 1 \\ g(x^{(2)}) = g(x^{(3)}) = -1 \end{array}\right\} \Leftrightarrow \left\{\begin{array}{l} g(x^{(1)}) = w_1 \cdot x_1^{(1)} + w_2 \cdot x_2^{(1)} + b = 1 \\ \Leftrightarrow 0 \cdot x_1 + (-1)\cdot(-1) + b = 1 \\ \Leftrightarrow b = \emptyset \\ g(x^{(2)}) = w_1 \cdot x_2^{(1)} + w_2 \cdot x_2^{(2)} + b = -1 \\ = \emptyset + (-1)\cdot 1 + b = -1 \Leftrightarrow b = \emptyset \\ g(x^{(3)}) = w_1 \cdot x_3^{(1)} + w_2 \cdot x_3^{(2)} + b = -1 \\ = \emptyset + (-1)\cdot 1 + b = -1 \Leftrightarrow b = \emptyset \end{array}\right.$$

∘ Per el que obtenim

$w = \begin{pmatrix}0\\-1\end{pmatrix}$ i $b = \emptyset$

Ademes del classificador

$$g\left(\begin{pmatrix}x_1\\x_2\end{pmatrix}\right) = 0 \cdot x_1 + (-1)\cdot x_2 = -x_2$$



e) Podem calcular el Margin de la següent manera:

$$\text{margin} = \frac{2}{\|w\|} = \frac{2}{\sqrt{0^2 + (-1)^2}} = \frac{2}{1} = 2$$

Alex Montoya Pérez (242873)

LAB 8: Ex 3 (Logistic Regression)

a) Amb la Regressió Logística multiclasse tindrem k funcions lineals seguides per el soft-max, tal que:

$$h_k(x) = \text{softmax}(w_k^T \cdot x + w_{k0})$$
terme biaix

· Assignarem el punt x a la classe C_k si $h_k(x) > h_l(x)$ per tot $l \neq k$

b) $w_1 = (1\ \emptyset\ -1)^T$ $\quad x^{(1)} = (1, 2)^T$
$w_2 = (\emptyset\ 1\ \emptyset)^T$
$w_3 = (\emptyset\ \emptyset\ 1)^T$

· Tindrem $h(x^{(1)}) = \dfrac{\exp(w_k^T \cdot x)}{\sum_{j=1}^{k} \exp(w_j^T \cdot x)}$, comencem calculant $w_k^T \cdot x$

$w_k^T \cdot x = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix}$, continuem fent l'exponencial

$\exp(w_k^T \cdot x) = \begin{pmatrix} e^{-1} \\ e \\ e^2 \end{pmatrix} = \begin{pmatrix} 1/e \\ e \\ e^2 \end{pmatrix}$, a continuació fem $\sum_{j=1}^{k} \exp(w_j^T \cdot x)$

$\sum_{j=1}^{k} \exp(w_j^T \cdot x) = \frac{1}{e} + e + e^2 = 10{,}48$, un cop tenim tots els paràmetres per separat ja podem calcular $h(x^{(1)}) = \dfrac{(1/e\ \ e\ \ e^2)}{10{,}48} = \begin{pmatrix} 0{,}04 \\ 0{,}26 \\ 0{,}70 \end{pmatrix}$

podem veure que la suma de les probabilitats ens dona 1, i que la probabilitat més gran es que $x^{(1)} \in C_3$ amb un 70%

SUPERFICIES DISCRIMINANTS (DS)

→ Sigui $g_k(x) = w_k^T \cdot x$, tindrem 3 DS

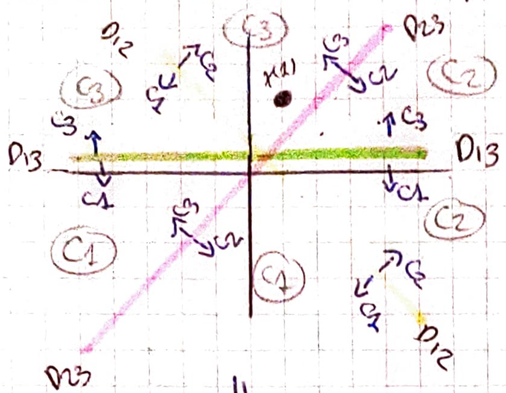
$$\begin{cases} g_1(x) = (1\ \emptyset\ 1)\begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix} = 1 - x_2 \\ g_2(x) = (\emptyset\ 1\ \emptyset)\begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix} = x_1 \\ g_3(x) = (\emptyset\ \emptyset\ 1)\begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix} = x_2 \end{cases}$$

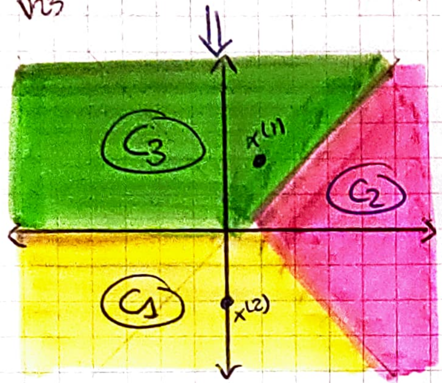
$DS_{12} \Rightarrow \{x \mid g_1(x) = g_2(x)\} = \{x \mid 1 - x_2 = x_1\}$
$= \{x \mid -1 + x_1 + x_2 = \emptyset\}$
$\Downarrow$
$x_2 = -x_1 + 1$

$DS_{13} \Rightarrow \{x \mid 1 - x_2 = x_2\} = \{x \mid 1 - 2x_2 = \emptyset\} \rightarrow x_2 = 1/2$

$DS_{23} \Rightarrow \{x \mid x_1 = x_2\} = \{x \mid x_1 - x_2 = \emptyset\} \rightarrow x_2 = x_1$

Dibuixem cada una de les superfícies discriminants





Δ

-> Observant D_{13} i el punt $(0,0)$ veiem:
$P(C_1) > P(C_3)$ ja que $1 - x_2 > x_2$

-> Observant D_{23} i el punt $(-1,1)$ veiem:
$P(C_2) < P(C_3)$ ja que $x_1 < x_2$

-> Observant D_{12} i el punt $(2,1)$ veiem:
$P(C_2) > P(C_1)$ ja que $x_1 > 1 - x_2$

=> · Podem veure que el punt $x^{(1)} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ pertany a la classe 3, mentres que el punt $x^{(2)} = \begin{pmatrix} 0 \\ -2 \end{pmatrix}$ pertany a la classe 1

?

El rol de les superfícies discriminants es delimitar cada una de les classes, el que ens permet assignar cada un dels punts a la classe que li pertoca

c) La relació entre la funció logística i la funció softmax es que la softmax generalitza la funció logística (utilitzant model lineal) ✱

"aixafa" K números (puntuació) funció exponencial. Proporciona K outputs normalitzats entre el ∅ i 1

Les sortides es poden interpretar com probabilitats a posteriori

$$p(C_k|x) = \frac{p(x|C_k) \cdot p(C_k)}{\sum_{j=1}^{k} p(x|C_j) \cdot p(C_j)} = \frac{\exp(a_k)}{\sum_{j=1}^{k} \exp(a_j)}$$

✱ model lineal $\Rightarrow h_w(x^{(n)}) = p(y_k = 1|x) = \frac{\exp(w_k^T x)}{\sum_{j=1}^{k} \exp(w_j^T \cdot x)}$ el qual requereix de K vectors $w_1 \dots w_k$ que podem representar com una matriu W de mida $(D+1) \cdot K$

✓

Δ $DS\{a\,b\}$ el podem determinar calculant $p(Ca) = p(Cb)$, per tant $p(Ca) - p(Cb) = \emptyset$ tindrem el següent:
- COSTAT POSITIU RECTA: on $r(x_1, x_2) > 0$, s'en $r(\cdot)$ l'equació de la recta, $p(Ca) - p(Cb) > 0$ per tant $p(Ca) > p(Cb)$
- COSTAT NEGATIU RECTA: $p(Ca) - p(Cb) < \emptyset$ per tant tenim $p(Ca) < p(Cb)$