Algoritmo Perceptrón[1]

Jorge Civera
Alfons Juan
Albert Sanchis

Departamento de Sistemas
Informáticos y Computación

[1]Para una correcta visualización, se requiere Acrobat Reader v. 7.0 o superior

Objetivos formativos

- Aplicar el algoritmo Perceptrón a una tarea de clasificación
- Explicar el comportamiento del algoritmo Perceptrón en función de sus parámetros

Índice

1. Funciones discriminantes lineales

Todo clasificador puede representarse como:

$$c(x) = \arg\max_{c} \; g_c(x)$$

donde cada clase c utiliza una ***función discriminante*** $g_c(x)$ que mide el grado de pertenencia de un objeto x a la clase c

Las funciones discriminantes más utilizadas son ***lineales*** (con $\boldsymbol{x}$):

$$\boxed{g_c(\boldsymbol{x}) = \boldsymbol{w}_c^t \boldsymbol{x} + w_{c0}} \quad \text{donde} \quad \boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_D \end{pmatrix} \quad \text{y} \quad \boldsymbol{w_c} = \begin{pmatrix} w_{c1} \\ \vdots \\ w_{cD} \end{pmatrix}$$

Con notación ***homogénea***:

$$g_c(\mathbf{x}) = \underline{\mathbf{w}_c^t \mathbf{x}} \quad \text{donde} \quad \mathbf{x} = \begin{pmatrix} 1 \\ \boldsymbol{x} \end{pmatrix} \quad \text{y} \quad \mathbf{w}_c = \begin{pmatrix} w_{c0} \\ \boldsymbol{w}_c \end{pmatrix}$$

$$g_c(x) = (1 \; x_1 \ldots x_D) \cdot \begin{pmatrix} w_{c0} \\ w_{c1} \\ \vdots \\ w_{cD} \end{pmatrix} = w_{c0} + w_{c1} \cdot x_1 + \ldots + w_{cD} \cdot x_D$$

2. Algoritmo Perceptrón

$\boxed{x_1}$ $c_1 = \text{spam}$
$\boxed{x_2}$ $c_2 = \text{spam}$
$\boxed{x_3}$ $c_3 = \text{no-spam}$
⋮ ⋮

Entrada: $\{(\mathbf{x}_n, c_n)\}_{n=1}^{N}$, $\boxed{\{\mathbf{w}_c\}_{c=0}^{C}}$, $\alpha \in \mathbb{R}^{>0}$ y $b \in \mathbb{R}$

Salida: $$\{\mathbf{w}_c\}^* = \arg\min_{\{\mathbf{w}_c\}} \sum_n \left[\max_{c \neq c_n} \mathbf{w}_c^t \mathbf{x}_n + b > \mathbf{w}_{c_n}^t \mathbf{x}_n \right]$$

Método: $$[P] = \begin{cases} 1 & \text{si} \quad P = \text{verdadero} \\ 0 & \text{si} \quad P = \text{falso} \end{cases}$$

Muestra bien clasificada

(x_n, c_n) $\boxed{g_{c_n}(x_n)} > g_c(x_n) \quad \forall c \neq c_n$

Función discriminante de la clase correcta x_n pertenece a c_n

repetir

para todo dato $\mathbf{x}_n$

$err = \texttt{falso}$

para toda clase c distinta de c_n

correcta

$g_c(x_n)$

si $\mathbf{w}_c^t \mathbf{x}_n + b > \mathbf{w}_{c_n}^t \mathbf{x}_n$: $\mathbf{w}_c = \mathbf{w}_c - \alpha \cdot \mathbf{x}_n$; $err = \texttt{verdadero}$

Incorrecta

$g_{c_n}(x_n)$ 0.1

si err: $\mathbf{w}_{c_n} = \mathbf{w}_{c_n} + \alpha \cdot \mathbf{x}_n$

vector de pesos de la clase incorrecta

hasta que no quedan muestras mal clasificadas

$\eta(x_n, c_n)\big|_{n=1}^{N}$

vector de pesos de la clase correcta

3. Ejemplo

$x_1 = \begin{pmatrix}0\\0\end{pmatrix}$ $\boxed{c_1 = 0}$ ← $g_0(x_1) > g_\bullet(x_1)$

$$\boxed{g_0(x_1) > g_\bullet(x_1)}$$

$$g_0(x_1) = w_0^t \cdot x_1 = (0\;0\;0)\cdot\begin{pmatrix}1\\0\\0\end{pmatrix} = 0\cdot 1 + 0\cdot 0 + 0\cdot 0 = 0$$

Funcion correcta

$$g_\bullet(x_1) = w_\bullet^t \cdot x_1 = (0\;0\;0)\begin{pmatrix}1\\0\\0\end{pmatrix} = 0$$ ← Func incorr

si $w_c^t \cdot x_1 + b > w_{c_1}^t \cdot x_1$

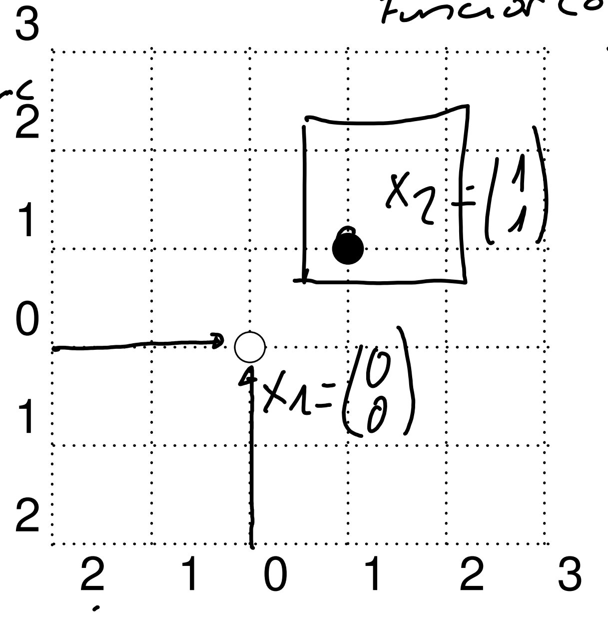
$$w_\bullet^t \cdot x_1 + b > w_0^t \cdot x_1$$

$$g_\bullet(x_1) + \overset{0.1}{b} > g_0(x_1)$$

si se cumple modificamos los pesos

$$w_\bullet = w_\bullet - \alpha \cdot x_1 = \begin{pmatrix}0\\0\\0\end{pmatrix} - 1.0 \cdot \begin{pmatrix}1\\0\\0\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix} - \begin{pmatrix}1\\0\\0\end{pmatrix} = \begin{pmatrix}-1\\0\\0\end{pmatrix}$$

$$w_0 = w_0 + \alpha \cdot x_1 = \begin{pmatrix}0\\0\\0\end{pmatrix} + 1.0 \cdot \begin{pmatrix}1\\0\\0\end{pmatrix} = \begin{pmatrix}1.0\\0\\0\end{pmatrix}$$ ⟸



$x_1 = \begin{pmatrix}0\\0\end{pmatrix}$ $c_1 = 0$ $w_0 = \begin{pmatrix}0\\0\\0\end{pmatrix}$ $\alpha = 1.0$

$x_2 = \begin{pmatrix}1\\1\end{pmatrix}$ $c_2 = \bullet$ $\boxed{w_\bullet = \begin{pmatrix}0\\0\\0\end{pmatrix}}$ $b = 0.1$

$x_2 = \begin{pmatrix}1\\1\\1\end{pmatrix}$ $c_2 = \bullet$ ← clase correcta para la muestra x_2 es $\bullet$

si $g_0(x_2) + b > g_\bullet(x_2)$ ent. modificamos pesos

$$g_0(x_2) = w_0^t \cdot x_2 = (1\;0\;0)\begin{pmatrix}1\\1\\1\end{pmatrix} = 1 + b$$

$$g_\bullet(x_2) = w_\bullet^t \cdot x_2 = (-1\;0\;0)\begin{pmatrix}1\\1\\1\end{pmatrix} = -1$$

$$\text{si } 1 + \overset{0.1}{b} > -1$$

$$w_0 = w_0 - \alpha \cdot x_2 = \begin{pmatrix}1\\0\\0\end{pmatrix} - 1.0\begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}0\\-1\\-1\end{pmatrix}$$

clase inco

$$w_\bullet = w_\bullet + \alpha \cdot x_2 = \begin{pmatrix}-1\\0\\0\end{pmatrix} + 1.0 \cdot \begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}0\\1\\1\end{pmatrix}$$

4. Convergencia y calidad de la solución

Converge si los datos son linealmente separables y $b \leq 0$

Conviene implementarlo con un máximo número de iteraciones.

Cuando $\alpha \to 0$, la convergencia es más suave, pero más lenta.

Calidad de la solución:

Linealmente separables	$b \leq 0$	$b > 0$
SI	Fronteras con poca holgura	Fronteras *centradas*
NO	Fronteras baja calidad	Fronteras casi óptimas

5. Conclusiones

Hemos visto:

- El algoritmo Perceptrón y una traza del mismo
- La convergencia del algoritmo en función de sus parámetros y las muestras de entrenamiento utilizadas

$$g_0(x) = \underset{w^t}{(1\ -1\ -1)} \cdot \begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix} = \underline{\underline{1 - x_1 - x_2}}$$

$$g_\bullet(x) = (-1\ 1\ 1) \cdot \begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix} = -1 + x_1 + x_2$$

$$g_0(x) = g_\bullet(x) \longrightarrow x_2 = \ldots$$