

Regresión Logística

Rafael Zambrano

Regresión Logística

- Utilizada para predecir probabilidades
- Ejemplo: Probabilidad de sufrir hipertensión

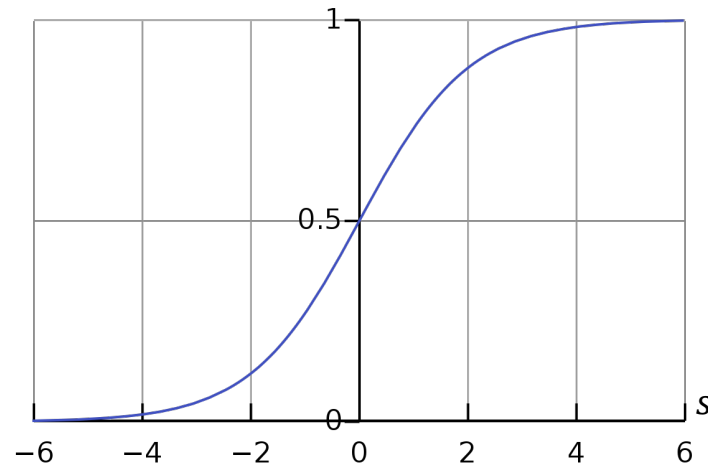
Edad	Altura	Peso	Colesterol	Fumador/a	Padece hipertensión
53	170	70	120	1	SÍ
67	156	85	240	0	NO
21	191	56	100	0	NO
34	182	77	500	1	SÍ
...

- Para un nuevo paciente, el modelo devolverá un valor de probabilidad de padecer hipertensión (entre 0 y 1)

Regresión Logística

- En la regresión lineal, teníamos $h(x) = \sum w_i x_i = \mathbf{w}^T \mathbf{x} = w_0 + w_1 x_1 + w_2 x_2 + \dots$
- En regresión logística, tenemos $h(x) = \theta(\sum w_i x_i) = \theta(\mathbf{w}^T \mathbf{x})$
- Función sigmoide

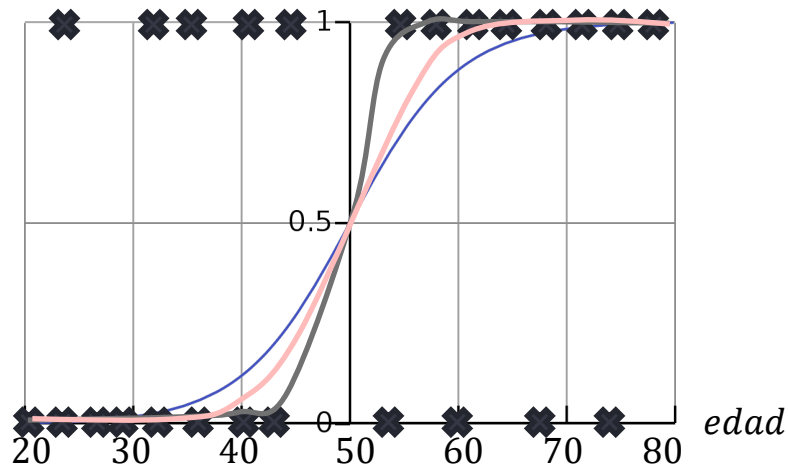
$$\theta(s) = \frac{e^s}{1+e^s}$$



Regresión Logística

- A modo de ejemplo, vamos a usar solamente la edad como variable explicativa de padecer hipertensión
- $h(x) = \theta(w_0 + w_1 \cdot edad)$

Edad	Padece hipertensión
53	1
67	0
21	0
34	1
...	...



Regresión Logística

- El hecho de padecer hipertensión se ve afectado por una probabilidad a la cual no tenemos acceso ni podemos medir
- Sólo podemos observar la ocurrencia de un evento e intentar **inferir** esa probabilidad

$$P(y|x) = \begin{cases} \theta(\mathbf{w}^T \mathbf{x}) & \text{si } y = +1 \\ 1 - \theta(\mathbf{w}^T \mathbf{x}) & \text{si } y = -1 \end{cases} = \begin{cases} \theta(w_0 + w_1 \cdot edad) & \text{si } y = +1 \\ 1 - \theta(w_0 + w_1 \cdot edad) & \text{si } y = -1 \end{cases}$$

$$\begin{aligned} w_0 &= -3.2 \\ w_1 &= 0.07 \end{aligned}$$

Edad	Padece hipertensión	$\theta(w_0 + w_1 \cdot edad)$	P(y x)
53	1	0.62	0.62
67	-1	0.82	0.18
21	-1	0.15	0.85
34	1	0.31	0.31
...

Hay que escoger los pesos w tales que maximicen el producto de las $P(y|x)$

$$\Rightarrow \max \prod P(y|x)$$

$$\prod P(y|x) = 0.62 \times 0.18 \times 0.85 \times 0.31 \times \dots$$

Regresión Logística

- Hay que escoger los pesos w tales que maximicen el producto de las $P(y|x)$

$$P(y|x) = \begin{cases} \theta(\mathbf{w}^T \mathbf{x}) & \text{si } y = +1 \\ 1 - \theta(\mathbf{w}^T \mathbf{x}) & \text{si } y = -1 \end{cases} = \begin{cases} \theta(\mathbf{w}^T \mathbf{x}) & \text{si } y = +1 \\ \theta(-\mathbf{w}^T \mathbf{x}) & \text{si } y = -1 \end{cases}$$

$$P(y|x) = \theta(y \cdot \mathbf{w}^T \mathbf{x})$$

$$\Rightarrow \text{maximizar } \prod \theta(y \cdot \mathbf{w}^T \mathbf{x})$$

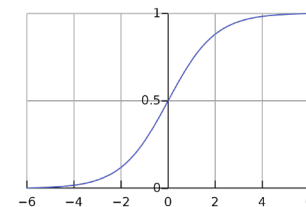
$$\Rightarrow \text{maximizar } \frac{1}{N} \ln \prod \theta(y \cdot \mathbf{w}^T \mathbf{x})$$

$$\Rightarrow \text{minimizar } -\frac{1}{N} \ln \prod \theta(y \cdot \mathbf{w}^T \mathbf{x})$$

$$\Rightarrow \text{minimizar } \frac{1}{N} \sum -\ln \theta(y \cdot \mathbf{w}^T \mathbf{x})$$

$$\Rightarrow \text{minimizar } \frac{1}{N} \sum \ln \frac{1}{\theta(y \cdot \mathbf{w}^T \mathbf{x})}$$

$$\Rightarrow \text{minimizar } \frac{1}{N} \sum \ln (1 + e^{-y \cdot \mathbf{w}^T \mathbf{x}})$$



$$1 - \theta(s) = \theta(-s)$$

$$\theta(s) = \frac{e^s}{1+e^s} = \frac{1}{1+e^{-s}} \\ \Rightarrow \frac{1}{\theta(s)} = 1 + e^{-s}$$

Regresión Logística

- Minimizar $E_{in}(\mathbf{w}) = \frac{1}{N} \sum \ln(1 + e^{-y\mathbf{w}^T x})$
- A diferencia de la regresión lineal, no hay una solución analítica para esta minimización
- La función $E_{in}(\mathbf{w})$ sólo tiene un mínimo
- Para minimizar la función se utiliza el método del **Gradiente Descendente**

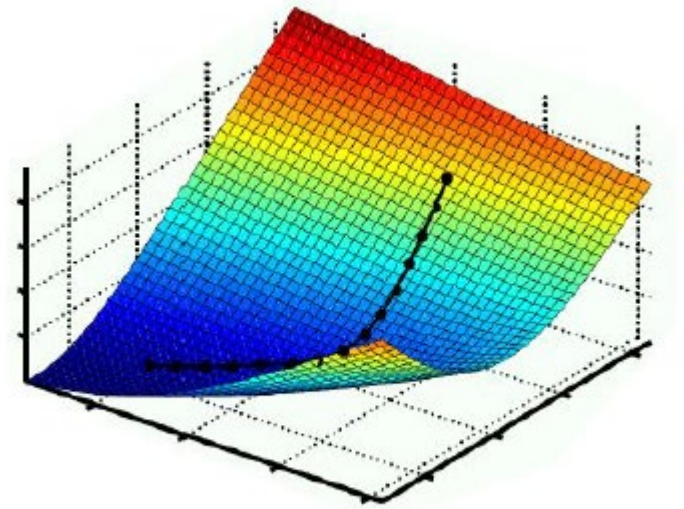


Regresión Logística

- Minimizar $E_{in}(\mathbf{w}) = \frac{1}{N} \sum \ln(1 + e^{-y\mathbf{w}^T \mathbf{x}})$
- La mejor dirección para “moverse” es el negativo del gradiente

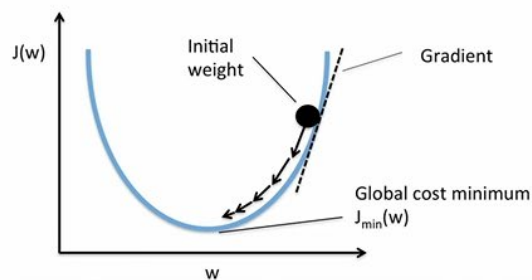
$$\mathbf{w}(t+1) = \mathbf{w}(t) - \alpha \nabla E_{in}$$

$$\nabla E_{in} = -\frac{1}{N} \sum \frac{y \cdot \mathbf{x}}{1 + e^{y\mathbf{w}^T \mathbf{x}}}$$

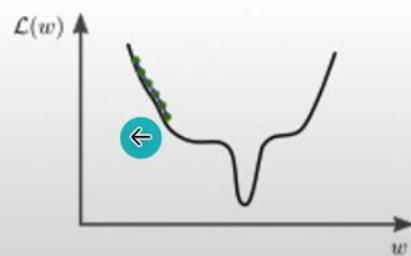


$\mathbf{w}(0) \rightarrow \mathbf{w}(1) \rightarrow \mathbf{w}(2) \rightarrow \mathbf{w}(3) \dots$

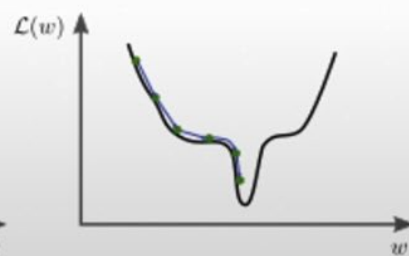
Regresión Logística



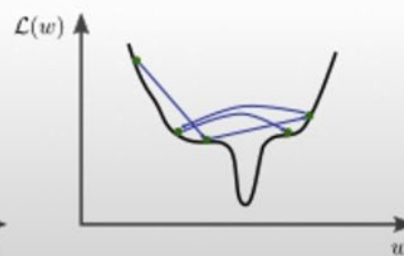
$$w = w - \alpha \cdot dw$$



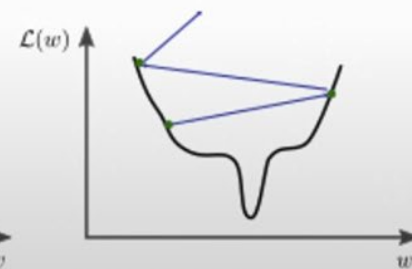
Learning rate too low



Good learning rate

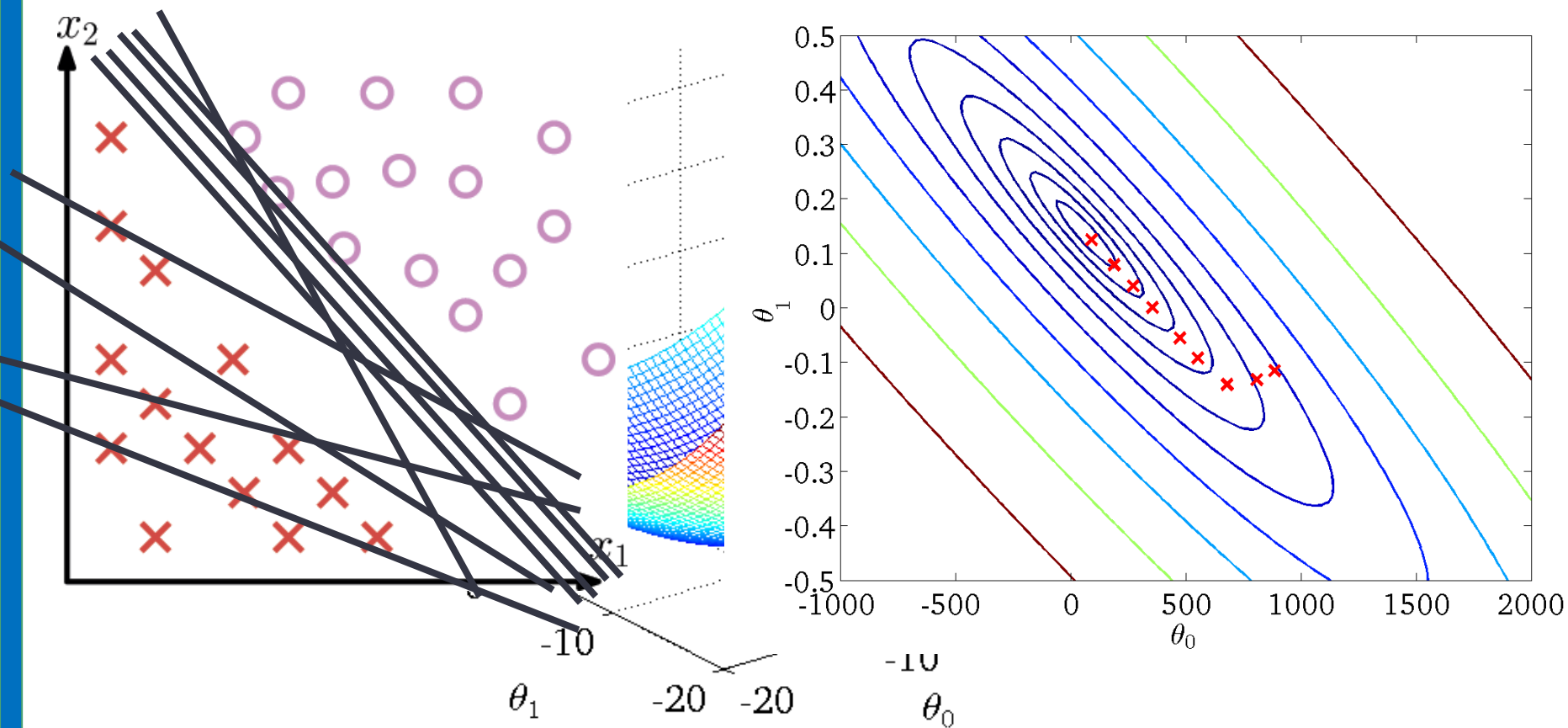


High learning rate



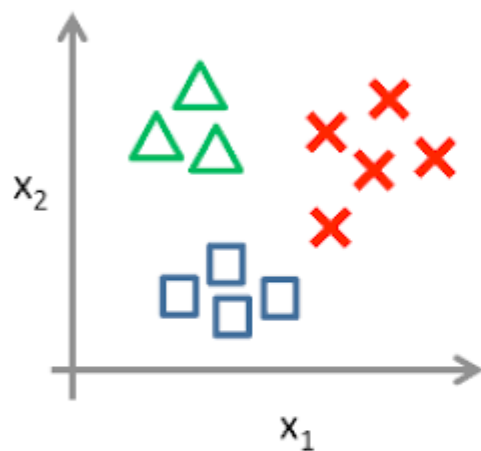
Learning rate much too high

Regresión Logística



Clasificación multiclase

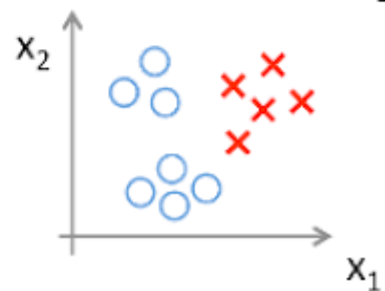
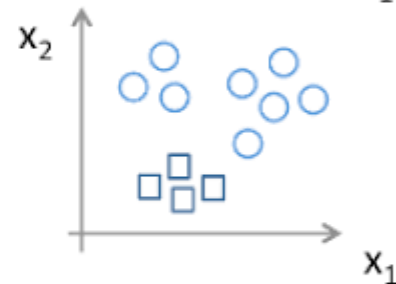
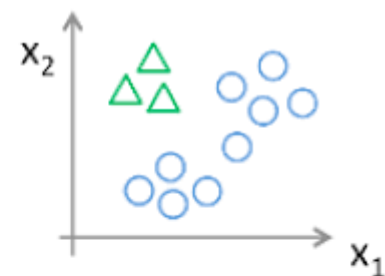
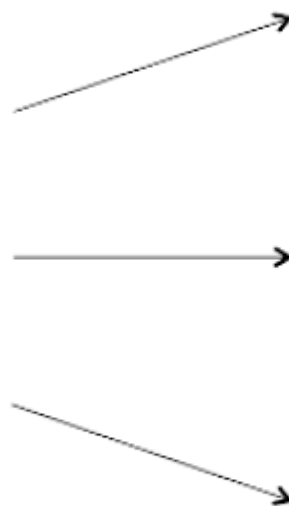
One-vs-all (one-vs-rest):



Class 1: Green

Class 2: Blue

Class 3: Red



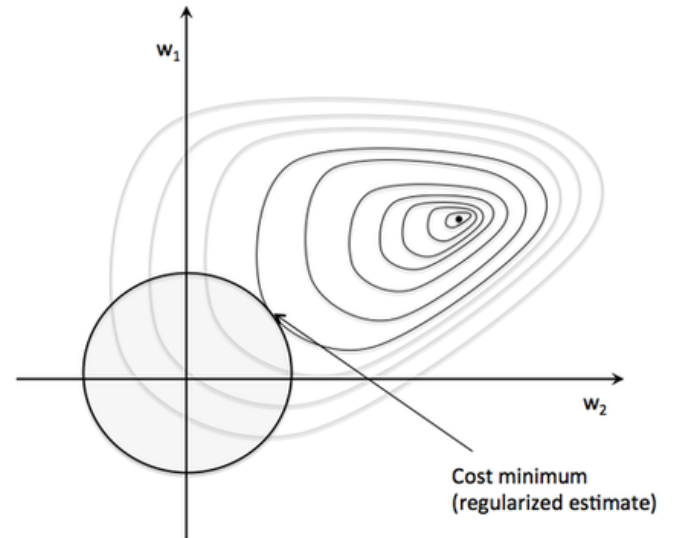
Regularización

- Regularización L2:

$$E_{in}(\mathbf{w}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum \ln(1 + e^{-y \mathbf{w}^T x})$$

- Regularización L1:

$$E_{in}(\mathbf{w}) = \|\mathbf{w}\|_1 + C \sum \ln(1 + e^{-y \mathbf{w}^T x})$$



Gracias.