Aprendizaje automatizado

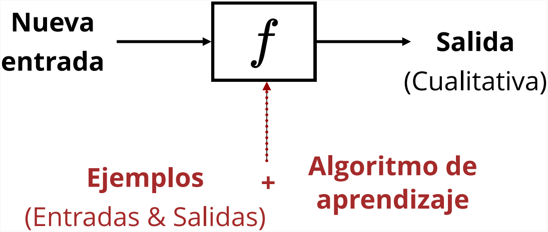
Regresión logística

Gibran Fuentes Pineda

Febrero 2023

Clasificación

- Salida discreta (cualitativa)
- Ejemplos: detección de spam, reconocimiento de rostros, etc.



Ejemplo de clasificación

- Clasificar sub-especias de la flor Iris basado en el ancho y largo de su pétalo

Ancho	Largo	Especie
1.4	0.2	Setosa
1.7	0.4	Setosa
1.5	0.1	Setosa
⋮	⋮	⋮
4.7	1.4	Versicolor
4.5	1.5	Versicolor
3.3	1.0	Versicolor
⋮	⋮	⋮

Características o atributo

Respuesta

Setosa

Versicolor

Tomada de `https://en.wikipedia.org/wiki/Iris_flower_data_set`

Clasificación: el caso binario

- En regresión lineal tenemos

$$P(y|\mathsf{x}, \boldsymbol{\theta}, \sigma^2) = \mathcal{N}(y|f_{\boldsymbol{\theta}}(\mathsf{x}, \boldsymbol{\theta}), \sigma^2)$$

- ¿Cómo podemos extender este modelo para la clasificación binaria?

Clasificación: el caso binario

- En regresión lineal tenemos

$$P(y|\mathrm{x}, \boldsymbol{\theta}, \sigma^2) = \mathcal{N}(y|f_{\boldsymbol{\theta}}(\mathrm{x}, \boldsymbol{\theta}), \sigma^2)$$

- ¿Cómo podemos extender este modelo para la clasificación binaria?

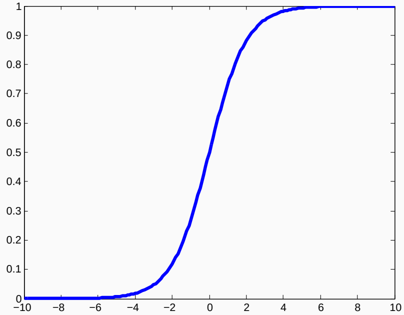
- Modelo de regresión logística

$$P(y|\mathrm{x}, \boldsymbol{\theta}) = Ber(y|sigm(\boldsymbol{\theta}^\top \mathrm{x}))$$

La función logística

- La función sigmoide o logística está dada por

$$sigm(z) = \frac{1}{1 + exp(-z)}$$



Estimador de máxima verosimilitud para regresión logística

- Tomando el negativo de la verosimilitud logarítmica

$$NVL(\boldsymbol{\theta}) = -\sum_{i=1}^{n}\{y^{(i)}\log q^{(i)} + (1-y^{(i)})\log(1-q^{(i)})\} = E(\boldsymbol{\theta})$$

donde $E(\boldsymbol{\theta})$ se conoce como *entropía cruzada binaria* y $q^{(i)} = sigm(\boldsymbol{\theta}^{\top}\mathrm{x}^{(i)})$

Estimador de máxima verosimilitud para regresión logística

- Tomando el negativo de la verosimilitud logarítmica

$$NVL(\boldsymbol{\theta}) = -\sum_{i=1}^{n}\{y^{(i)}\log q^{(i)} + (1 - y^{(i)})\log(1 - q^{(i)})\} = E(\boldsymbol{\theta})$$

  donde $E(\boldsymbol{\theta})$ se conoce como *entropía cruzada binaria* y $q^{(i)} = sigm(\boldsymbol{\theta}^{\top}\mathbf{x}^{(i)})$

- No hay solución cerrada, podemos entrenar usando descenso por gradiente

$$\nabla E(\boldsymbol{\theta}) = \sum_{i=1}^{n}(q^{(i)} - y^{(i)}) \cdot \mathbf{x}^{(i)} = \mathbf{X}^{\top}(\mathbf{q} - \mathbf{y})$$

Regularización en clasificación

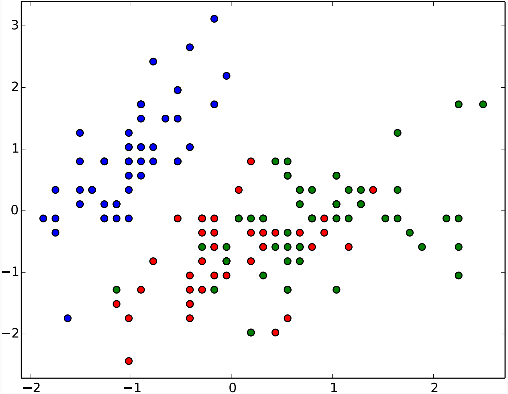
- Al igual que en regresión lineal la regularización puede ayudar a evitar el sobreajuste

- La función de error y el gradiente están dados por

$$\begin{aligned}\tilde{E}(\boldsymbol{\theta}) &= E(\boldsymbol{\theta}) + \lambda\|\boldsymbol{\theta}\|_2^2 \\ \nabla\tilde{E}(\boldsymbol{\theta}) &= \nabla E(\boldsymbol{\theta}) + 2\lambda\boldsymbol{\theta}\end{aligned}$$

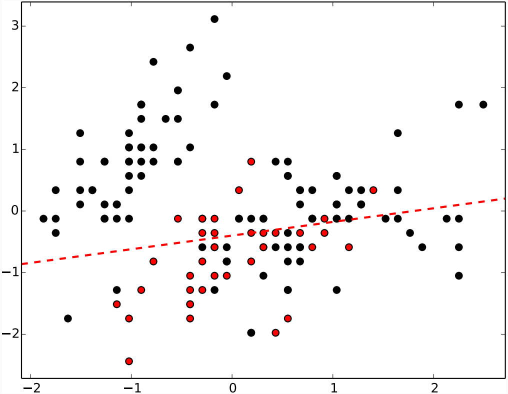
Clasificación multi-clase: uno vs el resto

- Un clasificador binario entre cada clase y el resto



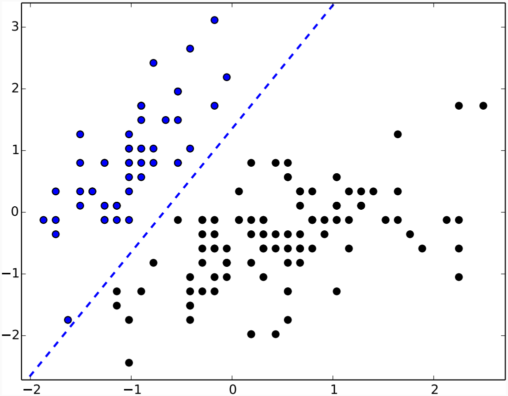
Clasificación multi-clase: uno vs el resto

- Un clasificador binario entre cada clase y el resto



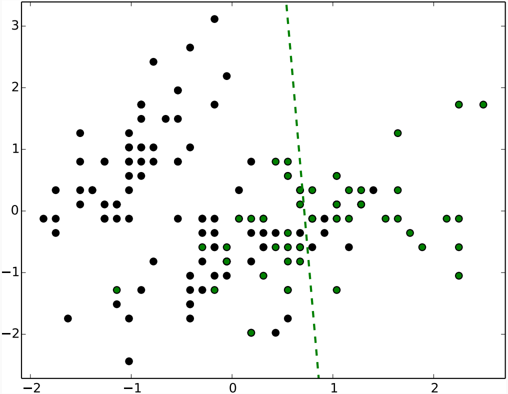
Clasificación multi-clase: uno vs el resto

- Un clasificador binario entre cada clase y el resto



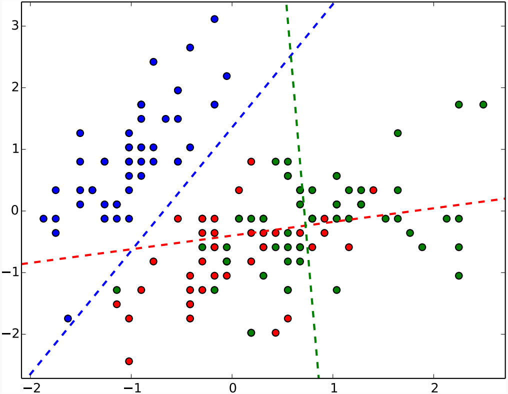
Clasificación multi-clase: uno vs el resto

- Un clasificador binario entre cada clase y el resto



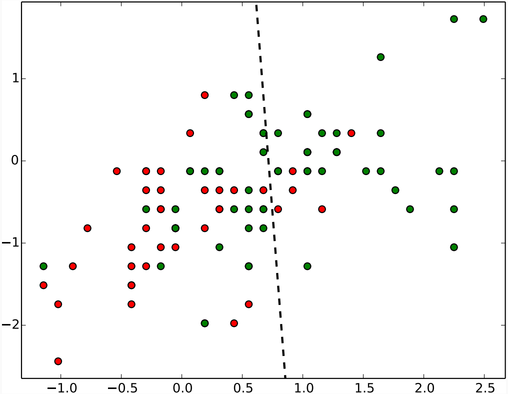
Clasificación multi-clase: uno vs el resto

- Un clasificador binario entre cada clase y el resto



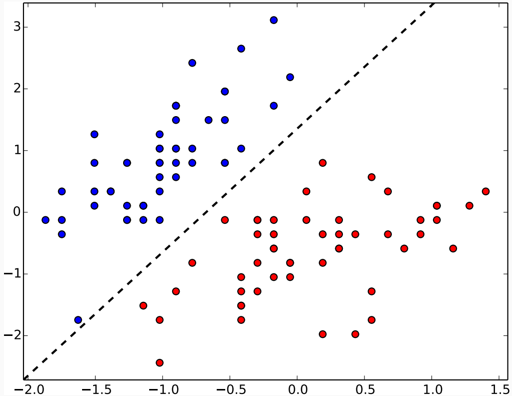
Clasificación multi-clase: uno vs uno

- Un clasificador binario entre cada par de clases



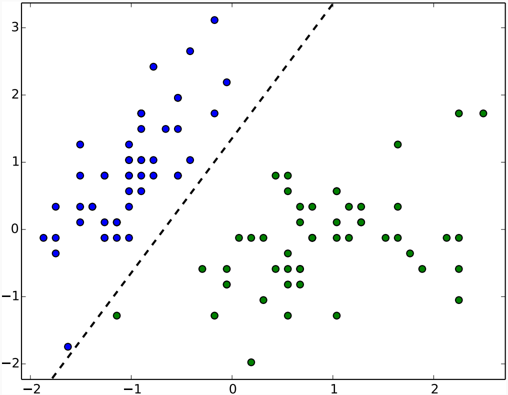
Clasificación multi-clase: uno vs uno

- Un clasificador binario entre cada par de clases



Clasificación multi-clase: uno vs uno

- Un clasificador binario entre cada par de clases



Clasificación multiclase: regresión logística multinomial

- Extensión de la regresión logística para múltiples clases

$$
\begin{aligned}
P(y|\mathbf{x}, \boldsymbol{\Theta}) &= Cat(y|softmax(\boldsymbol{\Theta}^\top \mathbf{x})_k) \\
&= \prod_{k=1}^{K} softmax(\boldsymbol{\Theta}^\top \mathbf{x})_k^{[y=k]}
\end{aligned}
$$

Clasificación multiclase: regresión logística multinomial

- Extensión de la regresión logística para múltiples clases

$$P(y|\mathbf{x}, \boldsymbol{\Theta}) = Cat(y|softmax(\boldsymbol{\Theta}^\top \mathbf{x})_k)$$
$$= \prod_{k=1}^{K} softmax(\boldsymbol{\Theta}^\top \mathbf{x})_k^{[y=k]}$$

- donde $\mathbf{x} = [1, x_1, \ldots, x_d]$, $[y = k]$ son los corchetes de Iverson, $\boldsymbol{\Theta} \in \mathbb{R}^{d \times K}$, $\boldsymbol{\Theta}^\top \mathbf{x} \in \mathbb{R}^K$ y *softmax* es una generalización de la función logística

$$softmax(\mathbf{z})_k = \frac{e^{z_k}}{\sum_{j=1}^{K} e^{z_j}} = \frac{e^{z_k - max(\mathbf{z})}}{\sum_{j=1}^{K} e^{z_j - max(\mathbf{z})}}$$

EMV para regresión logística multinomial

- Tomando el negativo de la verosimilitud logarítmica

$$NVL(\boldsymbol{\Theta}) = -\sum_{i=1}^{n}\sum_{k=1}^{K}[y^{(i)} = k] \log q_k^{(i)} = E(\boldsymbol{\Theta})$$

- donde

$$q_k^{(i)} = softmax(\boldsymbol{\Theta}^\top \mathbf{x}^{(i)})_k$$

- A $E(\boldsymbol{\Theta})$ se le como *entropía cruzada categórica.*

EMV para regresión logística multinomial

- Tomando el negativo de la verosimilitud logarítmica

$$NVL(\boldsymbol{\Theta}) = -\sum_{i=1}^{n}\sum_{k=1}^{K}[y^{(i)} = k] \log q_k^{(i)} = E(\boldsymbol{\Theta})$$

- donde

$$q_k^{(i)} = softmax(\boldsymbol{\Theta}^\top \mathbf{x}^{(i)})_k$$

- A $E(\boldsymbol{\Theta})$ se le como *entropía cruzada categórica.*

- Podemos entrenar modelos usando descenso por gradiente

$$\nabla E(\boldsymbol{\theta})_k = \sum_{i=1}^{n}(q_k^{(i)} - [y^{(i)} = k]) \cdot \mathbf{x}^{(i)}$$

¿Cómo representamos múltiples clases?

- **Sólo un valor**: se representa por una variable discreta y que puede tomar los valores $1, \ldots, K$. Por ej. si tenemos 4 clases, representamos la clase 2 por $y = 2$

¿Cómo representamos múltiples clases?

- **Sólo un valor**: se representa por una variable discreta y que puede tomar los valores $1, \dots, K$. Por ej. si tenemos 4 clases, representamos la clase 2 por $y = 2$

- **1-de-K**: cada clase se representa por un vector binario $\mathbf{y}$ de K dimensiones con 1 sólo en la posición de la clase. Siguiendo el mismo ejemplo tenemos

$$\mathbf{y} = [0, 1, 0, 0]$$