Introducción a la Inteligencia Artificial

Facultad de Ingeniería

Universidad de Buenos Aires

Índice

Clase 6

1. Clasificación Binaria
    a. Motivación
    b. Regresión Logística - Ejercicio de Aplicación
    c. Regresión Logística - Teoría
2. Clasificación Multiclase
    a. Motivación
    b. Softmax
    c. Ejercicio de Aplicación
3. Ejercicio integrador

IA → ML

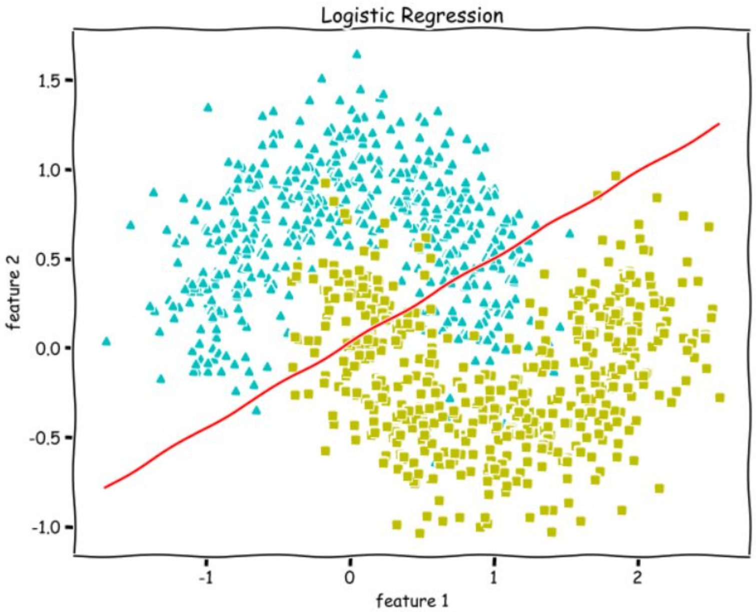
- Tareas a resolver
  - optimizacion | decision making
  - regresion $X \in \mathbb{R}^{n \times p}$, $y \in \mathbb{R}^n$
  - clasificacion $X \in \mathbb{R}^{n \times p}$, $y \in K$
- Métricas → Performance
  - overfitting
  - underfitting
- Preprocesamiento de datos → Adaptar la entrada para mejorar la performance
- Modelos
  - tipo de solución
    - cerrada (OLS)
    - numérica (perceptrón)
  - tipo de problema
    - supervisado
    - no supervisado
    - semi supervisado

Clasificación Binaria - Motivación

modelo lineales → reg. logistica, SVM Lineal, LDA

modelos no lineales → KNN, KSVM, CDA

- Credit Card Fraudulent Transaction Detection
- Medical Diagnosis
- Spam Detection
- Sentiment Analysis
- Binary Image Classification



Clasificación: Dado un $\bar{\bar{X}} \in \mathbb{R}^{n \times p}$ quiero mapear los x_{ij} a K clases (binario $K=2$).

. Clases separables: $\forall x_{ij} \in g \,/\, g \in K \;\; D(\gamma)=1$

cada input tiene una única clase ⇒ me dice que en el espacio de features existe una región (hiperplano, hiper región) que separa mis datos.

Métricas

MSE (gato, perro)? ⇝ en clasif. usamos la matriz de confusión.

real \ Pred	T	F
GL* T	TP	FN
GL* F	FP	TN

← prediction

$$y = \mathbb{I}_{c=k} = \begin{cases} 1 & \text{si } y_i = k \\ 0 & \text{si } y_i \neq k \end{cases}$$

* GL: Golden Label
Ground truth
Real values

- Accuracy $= \dfrac{TP + TN}{\#\,\text{muestras}} \in [0,1]$
- Specificity (TN rate) $= \dfrac{TN}{TN + FP}$
- precision $= \dfrac{TP}{TP + FP}$
- recall $= \dfrac{TP}{P}$
- $F_1 = 2 \cdot \dfrac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = \dfrac{TP}{TP + \frac{1}{2}(FP + FN)}$
- $F_\beta = (1+\beta)^2 \dfrac{\text{prec.} \cdot \text{recall}}{\beta^2\,\text{prec} + \text{recall}}$ $(F_{0,25})$

Formas de clasificar:

+ Definimos un proceso de scoring ⇒ como resolver la clasificación.

$\uparrow$ complejidad ⇒ $\uparrow$ computo $\overset{?}{\Rightarrow}$ mejor modelo

① Modelo generativo: modelamos la distrib de inputs y outputs ⇒ generamos la salida (permite samplear y generar data sintética) (GAN's)

② Modelos discriminantes: planteo $\mathbb{P}(C_k \mid \bar{\bar{x}})$ ⇝ mediante inferencia bayesiana puedo obtener mi modelo.

③ Modelos de fn. discriminante: buscamos $f: \mathbb{R}^{n \times m} \longmapsto [C_1, \ldots, C_K]$

↳ Regresión logística: la idea es partir del modelo de OLS:

partimos de $y \sim \beta(1, \pi_i) \Rightarrow E(y) = \pi_i \,,\; Var(y) = \pi_i(1-\pi_i)$

me gustaría obtener algo como:

$$\pi_i = \bar{\bar{x}}_i^t \bar{\beta} \quad , \quad \beta \in \mathbb{R}^{p \times 1}$$

$\pi_i \in [0,1]$ pero $\bar{x}_i^t \bar{\beta}$ no lo hace $\Rightarrow$ tengo que transformar el problema

$$f : [0,1] \longmapsto \mathbb{R}_{>0} \quad \rightsquigarrow \quad odd_i = \frac{\pi_i}{1-\pi_i}$$

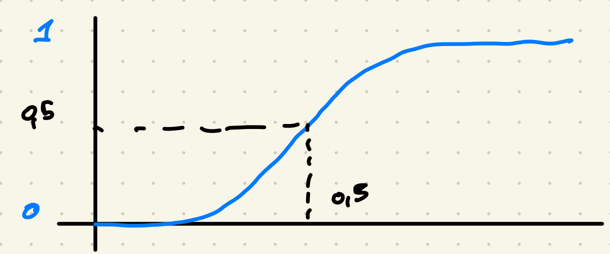
razon de probabilidad: es el cociente entre la probabilidad del evento A y su complemento.

Vamos a tomar el log de la razón de probabilidad, esto es una transf. biyectiva $\mathbb{R}_{>0} \longmapsto \mathbb{R}$. Esta transf. se la conoce como logit (log odds ratio):

$$\eta_i = \text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \bar{\bar{x}}_i^t \beta$$

existe: $\pi_i = \text{antilogit}(\eta_i) = \dfrac{e^{\eta_i}}{1+e^{\eta_i}}$

fn. sigmoide



Modelo de regresión logística

Sean $y_1, \ldots, y_n$ realizaciones de una binomial $y_i \sim \mathcal{B}(1, \pi_i)$.

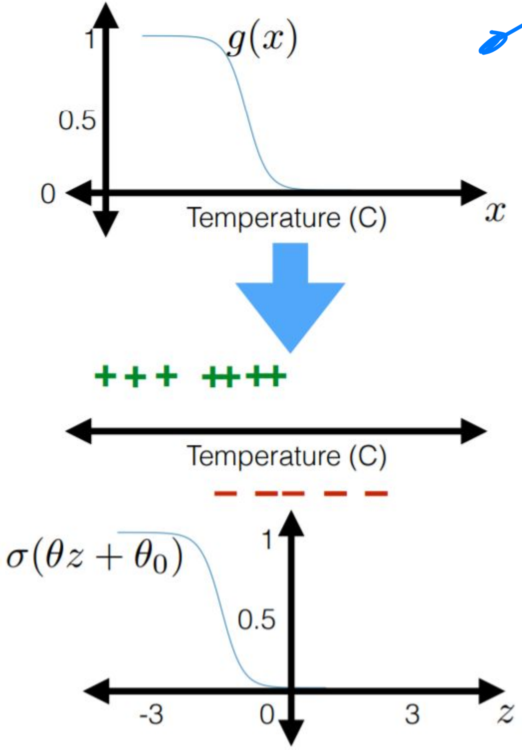
Asumimos la relación subyacente entre los datos x_i y el logit de la razon de probabilidad es lineal:

$$\text{logit}(\pi_i) = \beta_0 + \sum_{i=1}^{p} \beta_i \cdot x_i$$

el modelo queda definido como un modelo lineal generalizado con resp. binomial y fn. de enlace logit.

Clasificación Binaria - Motivación

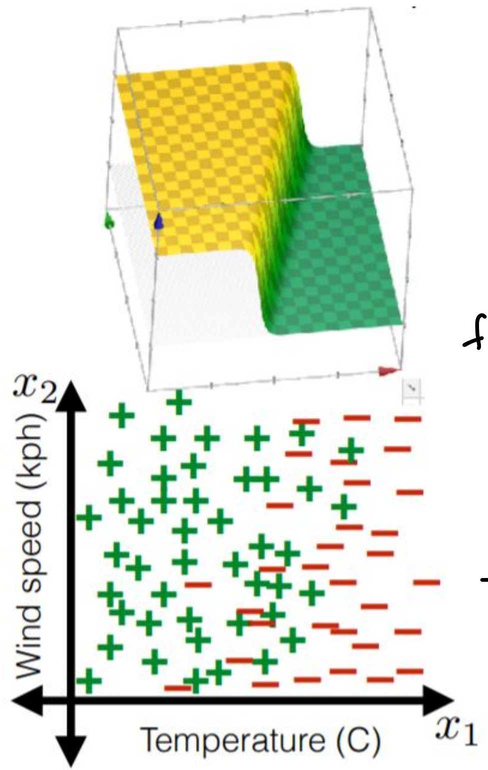
$(x,y) \to \boxed{RLC} \to c, p(c)$

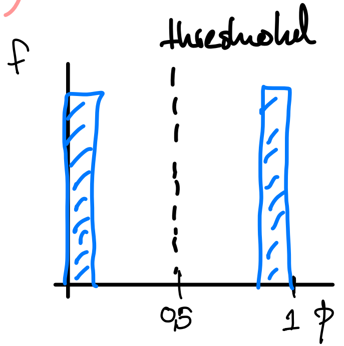


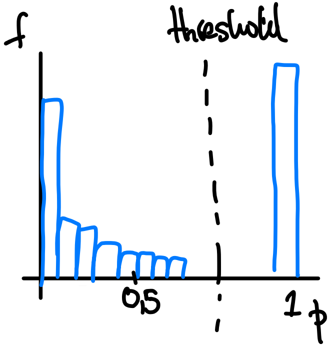
fn logística pertenece a la familia de fn. de squashing

$f: x, a \mapsto y$

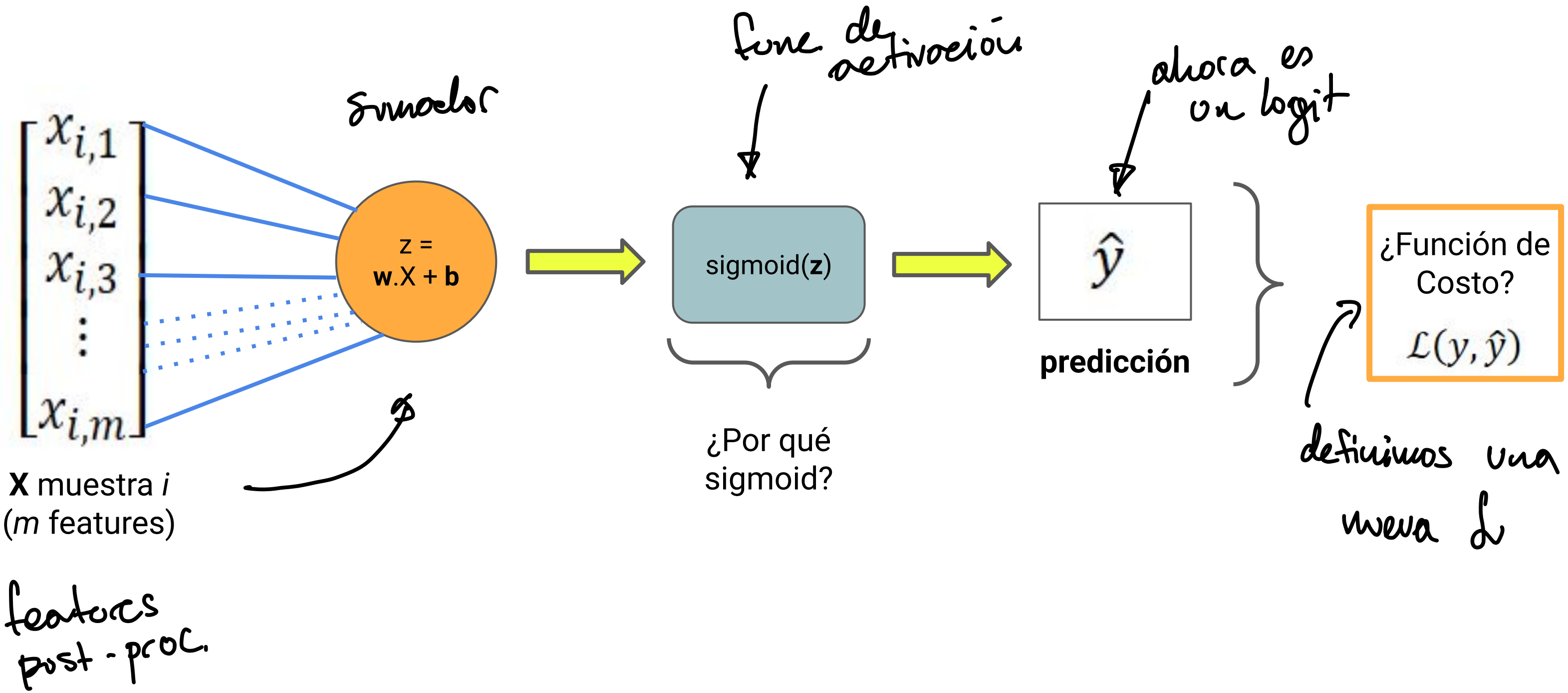
$\mathbb{R} \to [0,1]$







Regresión Logística



quiero mapear $\sigma(a)$, partiendo de $z = w^t x$

$$\sigma(z) = \sigma(w^t x) \rightsquigarrow \partial_a \sigma(a) = \sigma(a)(1-\sigma(a))$$ propiedad

$$\partial_w \sigma(z) = \sigma(w^t x)(1-\sigma(w^t x))\, x \quad ①$$

planteamos la verosimilitud: $\mathbb{P}(y|x) = \prod_{i=1}^{N} \hat{y}_i^{y_i} (1-\hat{y}_i)^{1-y_i} = \mathcal{L}$

$\log \mathcal{L} = \ell$

- N: cant. de muestras
- $\hat{y}$: predicción (prob)
- y: GL

$$\hat{y} = \mathbb{P}(c/x)$$

$$\ell = \sum_{i=1}^{N} \ln(\mathbb{P}_w(y_i = y_i / \bar{x}_i = x_i))$$ $\rightsquigarrow$ buscamos maximizar ℓ.

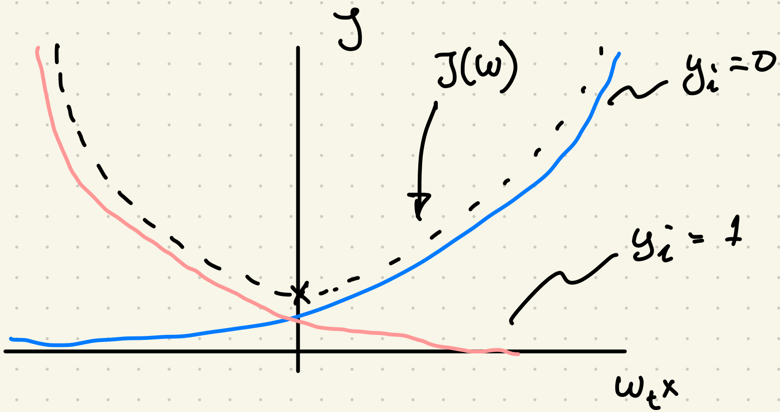
$$\max_w \sum_{i=1}^{N} \ln\left(\sigma(w^t x)^{y_i} \cdot (1-\sigma(w^t x))^{1-y_i}\right)$$

$$\max_w \sum_{i=1}^{N} y_i \ln(\sigma(w^t x)) + (1-y_i) \ln(1-\sigma(w^t x)) \rightsquigarrow x(-1)$$

$$\min_{w} \underbrace{\sum_{i=1}^{N} -y_i\, a - (1-y_i)\, b}_{1}$$

Vamos a aproximar 1 usando la binary crossentropy:

$$J(w) = \frac{1}{N} \cdot \sum_{i=1}^{N} \left[-y_i\, a - (1-y_i)\, b \right]$$



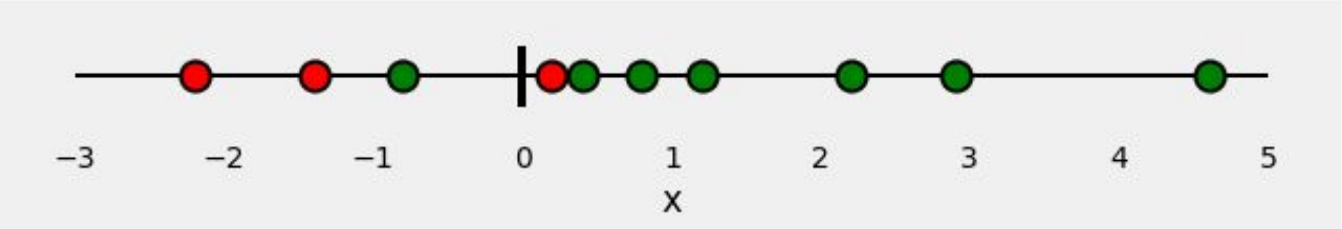
$$\nabla_w J(w) = 0$$

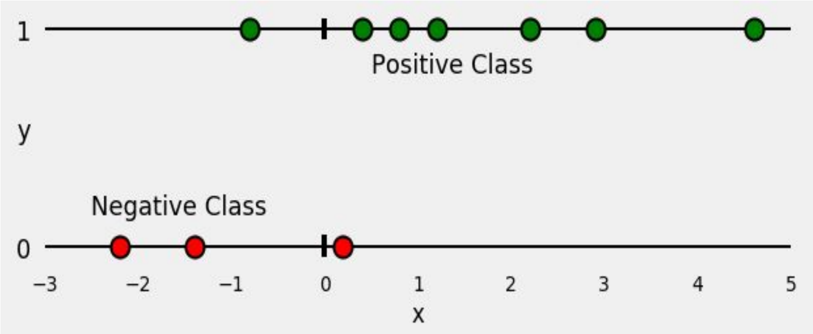
$$\nabla_w J(w) \to \text{GD}$$

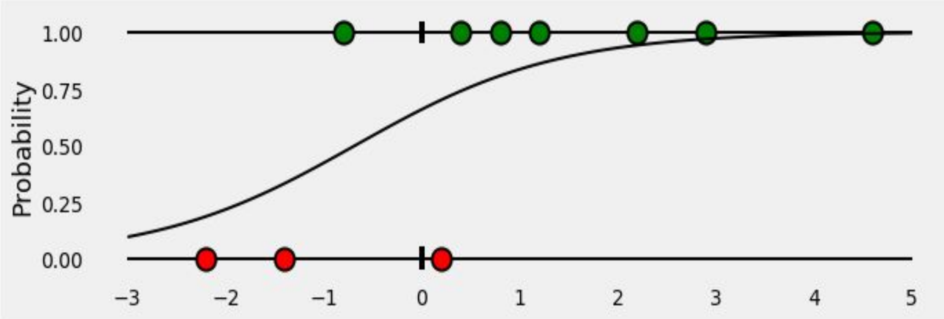
↳ SGD

↳ GDB

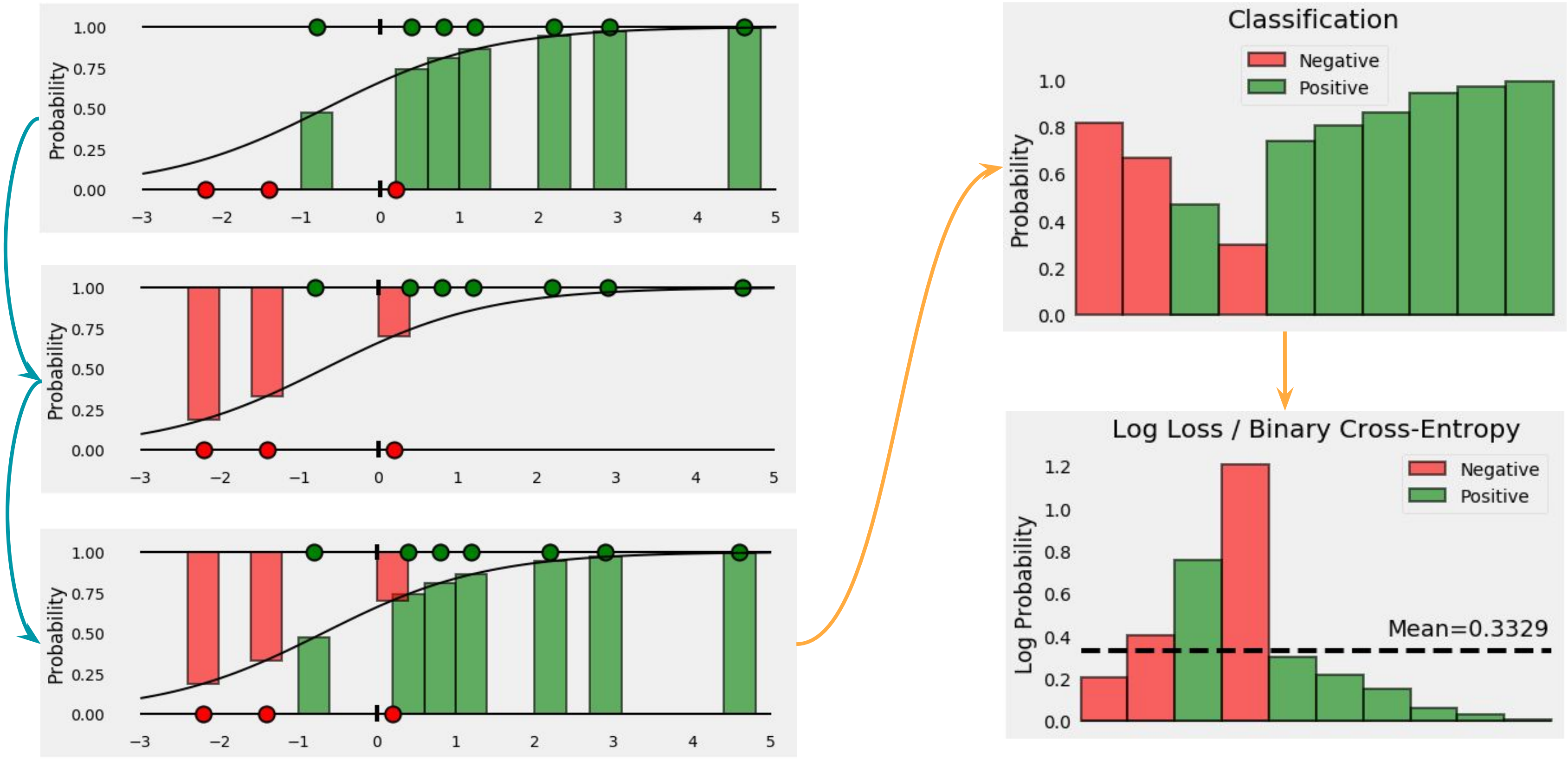
Regresión Logística





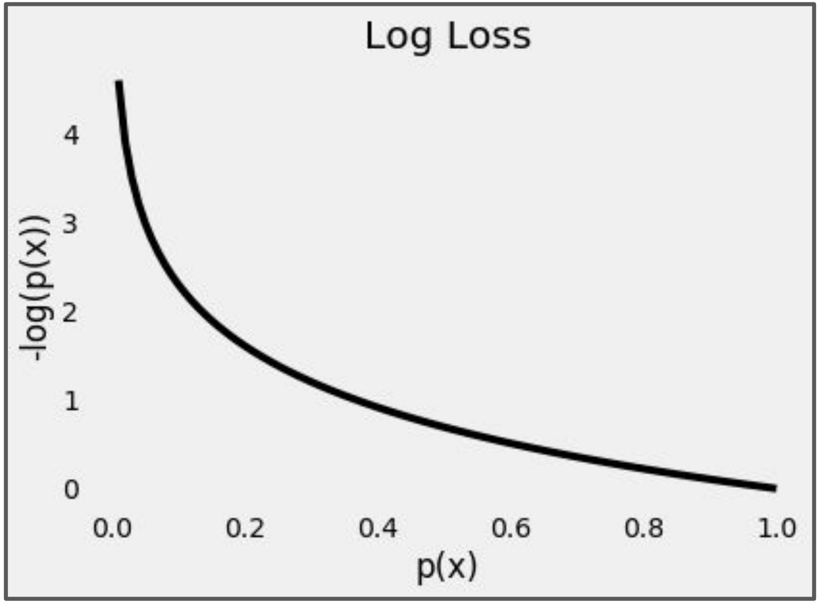


Regresión Logística



Regresión Logística

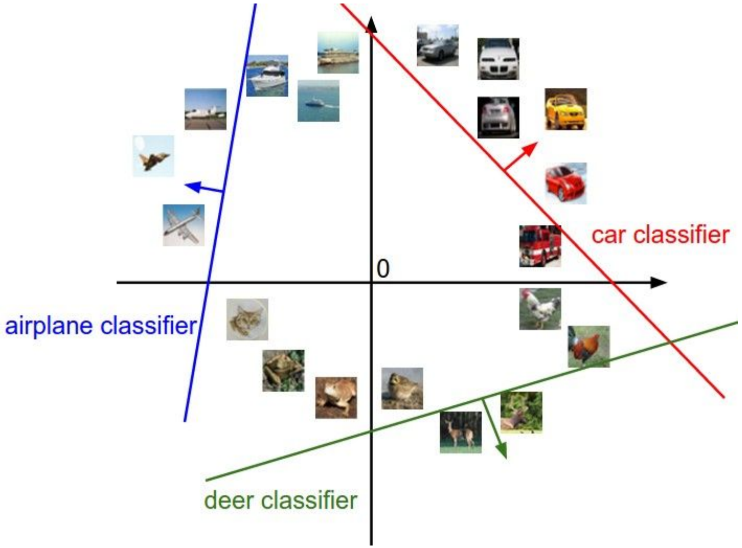
$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p(y_i)) + (1 - y_i) \cdot log(1 - p(y_i))$$

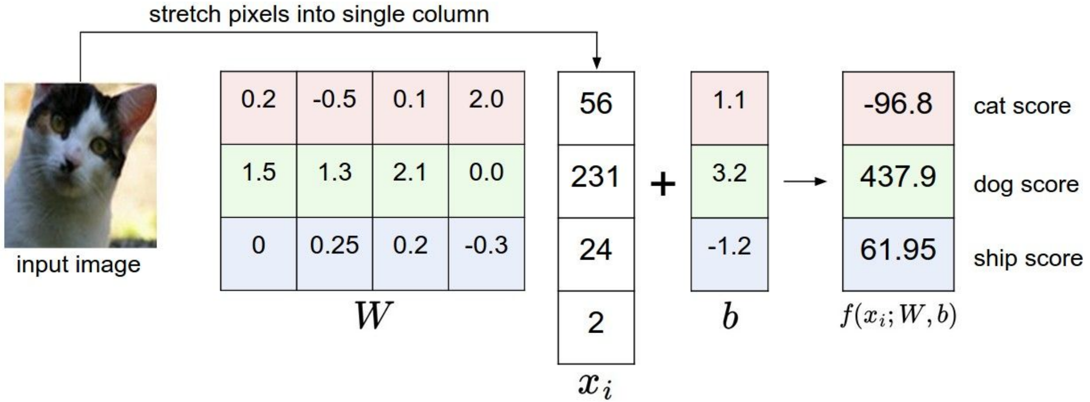


(1) REGRESIÓN LOGÍSTICA - EJERCICIO DE APLICACIÓN

(1) REGRESIÓN LOGÍSTICA - TEORÍA

Clasificación Multiclase - Motivación





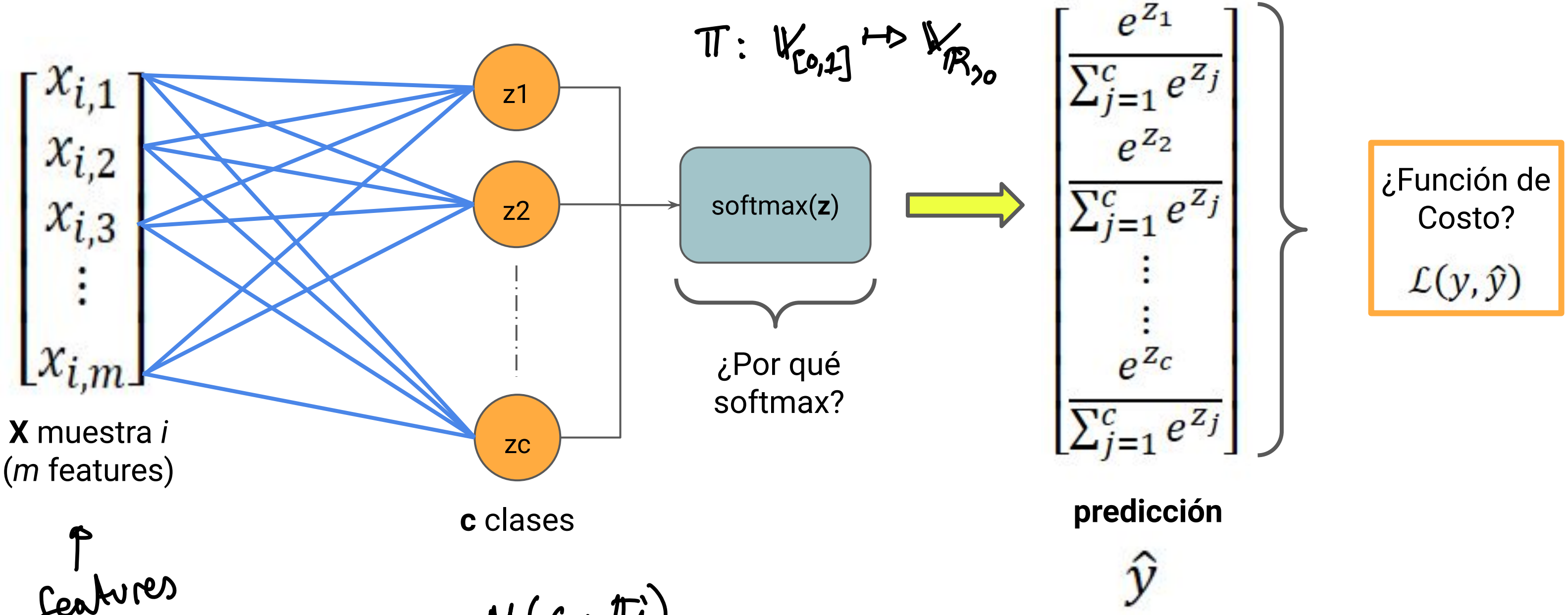
Softmax

c regresores

$z_i = w_i^t x + b_i$

$\pi : \mathbb{V}_{[0,1]} \mapsto \mathbb{V}_{\mathbb{R}_{>0}}$

$\mathbb{R}^{c \times 1}$



features pre proc.

$y_i \sim M(c, \pi_i)$

Softmax

$$P(y_1 | x_i ; w) = \frac{e^{w_1 x_i + b_1}}{\sum_j e^{z_i}}$$

$$f_{y_i} = z_i$$

$$P(y_i \mid x_i; W) = \frac{e^{f_{y_i}}}{\sum_j e^{f_j}}$$

$$\frac{e^{f_{y_i}}}{\sum_j e^{f_j}} = \frac{Ce^{f_{y_i}}}{C\sum_j e^{f_j}} = \frac{e^{f_{y_i} + \log C}}{\sum_j e^{f_j + \log C}}$$

q(x)

$$H(p,q) = -\sum_x p(x) \log q(x)$$

1 [Softmax Forma Gráfica]

2 [Softmax Visualización 3D]

Softmax

Softmax

Derivación Softmax

$$p_i = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}}$$

$$\frac{\partial p_i}{\partial z_k} = \frac{\partial \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}}}{\partial z_k}$$

$$\frac{\partial p_i}{\partial z_k} = p_i(\delta_{ik} - p_k) \qquad \delta ik = \begin{cases} 1, i = k \\ 0, i \neq k \end{cases}$$

Derivación Cross-Entropy

$$L = -\sum_i y_i log(p_i)$$

$$\frac{\partial L}{\partial z_i} = -\sum_j y_j \frac{\partial log(p_j)}{\partial z_i}$$

$$= -\sum_j y_j \frac{\partial log(p_j)}{\partial p_j} \times \frac{\partial p_j}{\partial z_i}$$

$$= -\sum_j y_j \frac{1}{p_j} \times \frac{\partial p_j}{\partial z_i}$$

$$\frac{\partial L}{\partial z_i} = -y_i(1 - p_i) - \sum_{j \neq i} y_j \frac{1}{p_j}(-p_j . p_i)$$

$$= -y_i(1 - p_i) + \sum_{j \neq i} y_j . p_i$$

$$= p_i \left(y_i + \sum_{j \neq i} y_j \right) - y_i$$

$$\frac{\partial L}{\partial z_i} = p_i - y_i$$

$$\frac{\partial z_i}{\partial W} = x_i$$

$$\frac{\partial L}{\partial W} = \sum_{i=1}^{N} (p_i - y_i)\, x_i$$

Usar gradiente descendente para actualizar W !!!

(1) SOFTMAX - EJERCICIO DE APLICACIÓN

Ejercicio integrador

1. Implementar el algoritmo de regresión logística en NumPy.
2. Aplicar el modelo a un dataset de elección.
3. Comparar los resultados con Scikit-Learn .
4. Comparar los resultados agregando regularización.

Bibliografía

- The Elements of Statistical Learning | Trevor Hastie | Springer
- An Introduction to Statistical Learning | Gareth James | Springer
- Deep Learning | Ian Goodfellow | https://www.deeplearningbook.org/
- Mathematics for Machine Learning | Deisenroth, Faisal, Ong
- Artificial Intelligence, A Modern Approach | Stuart J. Russell, Peter Norvig
- Understanding binary cross-entropy: a visual explanation | Daniel Godoy
- Visual Information Theory | Link
- https://cs231n.github.io/
- Classification and Loss Evaluation-Softmax and Cross Entropy Loss | Paras Dahal