Aprendizaje Automático y Análisis de Datos

Conocimiento de los datos

Julián Gil González

julian.gil@javerianacali.edu.co (Periodo 2023-I)

Agenda

Temas:

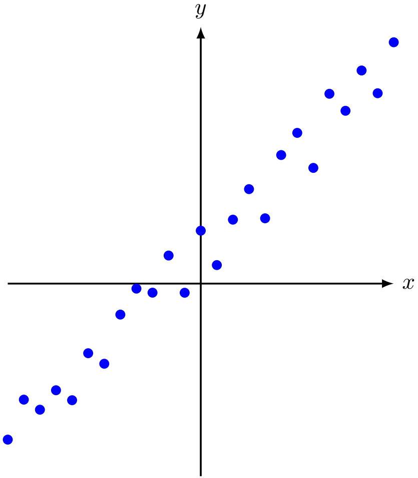
- Regresión lineal.
- Clasificación lineal binaria.
- Clasificación lineal multi-clase.

Objetivos del aprendizaje: Al final de esta clase los estudiantes estarán en la capacidad de:

- Plantear problemas de clasificación y regresión que se puedan solucionar a través de los métodos estudiados.
- Explicar qué los problemas de clasificación linealmente separables y cuál es la consecuencia de que lo sean.

Table of Contents

Modelo I

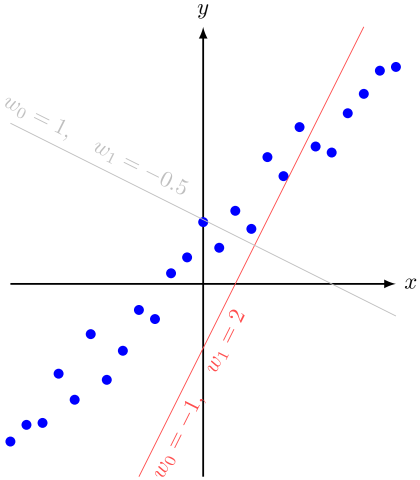


- Suponer el problema de regresión (de una dimensión) mostrado en la figura.
- El modelo de regresión lineal describe la **relación lineal** entre las entradas x y las salidas y. Así:

$$y = w_0 + w_1 * x,$$

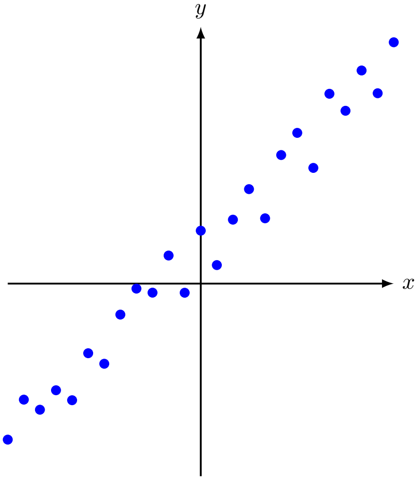
donde w_0 y w_1 respectivamente representan el intercepto y pendiente de la recta.

Modelo II



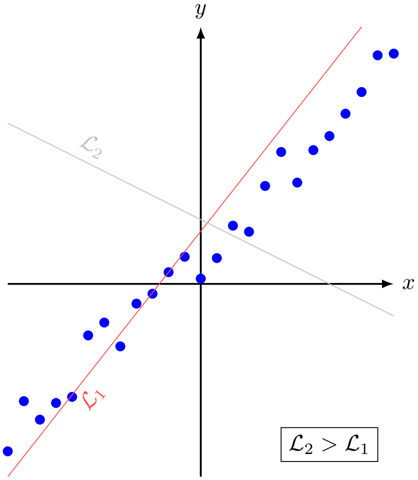
- Los parámetros w_0 y w_1 pueden tomar cualquier valor real.
- Diferentes valores de los parámetros w_0 y w_1 producen rectas distintas.

Función de costo o función de pérdida I



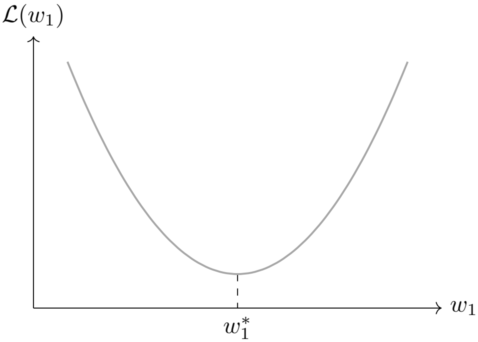
- Dada la nube de puntos $\{x^i, y^i\}$ mostrada en la Figura, es necesario calcular la mejor combinación de parámetros w_0 y w_1.
- La función de costo o de pérdida $\mathcal{L}(w_0, w_1)$ permite evaluar qué tan bien se ajusta el modelo (con parámetros w_0, w_1) a los datos.

Función de costo o función de pérdida II



- Suponer que se tienen las dos rectas mostradas en la Figura.
- A cada una se le calcula su función de pérdida.
- La recta que genera el menor valor en la función de costo es la que mejor se ajusta a los datos.

Función de costo o función de pérdida III

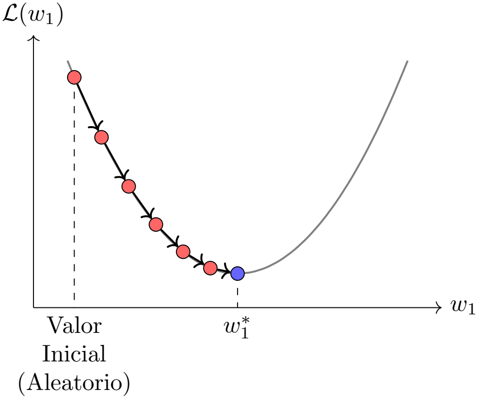


- Para facilidad en la exposición, se asume que w_0 se conoce; así $\mathcal{L}(w_1)$ corresponde a una función unidimensional.
- Para las tareas de regresión se suele usar el error cuadrático medio como función de costo. Así:

$$\mathcal{L}(w_1) = \sum_{i=1}^{N} (w_1 * x - y)^2$$

- Así, es necesario encontrar el valor de w_1 que minimice la función de pérdida.

Gradiente descendente I

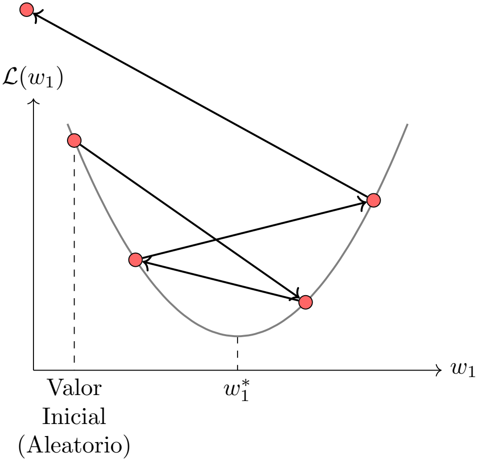


- Para encontrar el mínimo de la función de pérdida, se suele usar el algoritmo del gradiente descendente (o alguna de sus mejoras).
- Es un proceso iterativo donde el valor actual del parámetro depende del valor en la iteración pasada y del valor de la derivada.

$$w_1^{(t)} = w_1^{(t-1)} - \eta \frac{d\mathcal{L}(w_1)}{dw_1},$$

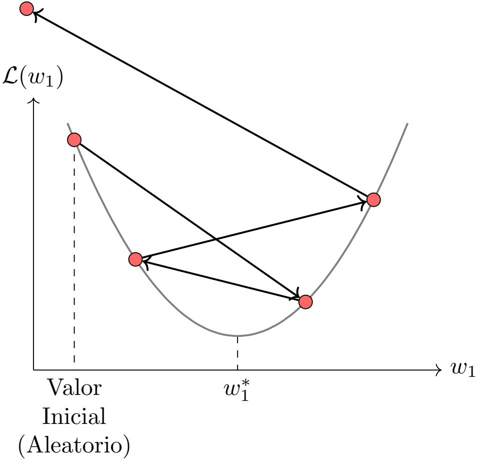
donde η es el factor de aprendizaje

Gradiente descendente II



- $\frac{d\mathcal{L}(w_1)}{dw_1}$, representa la mayor tasa de aumento de la función $\mathcal{L}(w_1)$.
- Así, $-\frac{d\mathcal{L}(w_1)}{dw_1}$ indica el “camino” por el cual $\mathcal{L}(w_1)$ decrece más rápido.
- El parámetro η controla el tamaño del paso. Si este no se elige bien, el algoritmo puede divergir.

Gradiente descendente III



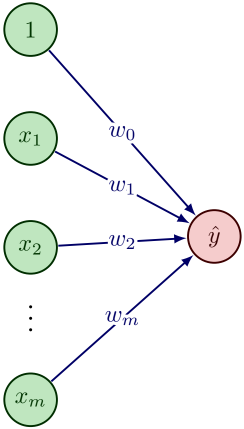
- El algoritmo del gradiente descendente también puede usarse para múltiples parámetros. Por ejemplo, volvamos al modelo original donde tenemos dos parámetros w_0 y w_1.
- En este caso el algoritmo se aplica individualmente a cada parámetro. Así,

$$
\begin{aligned}
w_0^{(t)} &= w_0^{(t-1)} - \eta \frac{\partial \mathcal{L}(w_0)}{\partial w_0}, \\
w_1^{(t)} &= w_1^{(t-1)} - \eta \frac{\partial \mathcal{L}(w_1)}{\partial w_1},
\end{aligned}
$$

Algoritmo de entrenamiento

input: Conjunto de entrenamiento $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n}$, α, Número de iteraciones T
output: parámetros estimados $\boldsymbol{w}$
Inicializar los parámetros $\boldsymbol{w}$ de forma aleatorias
for $t = 1$ *to* T **do**
  Calcular las predicciones $\hat{y}_i = \boldsymbol{w}^T \boldsymbol{x}_i$
  Calcular la función de costo $\mathcal{L}(\mathbf{w})$
  Calcular gradientes $\nabla_{\boldsymbol{w}} \mathcal{L}(\boldsymbol{w})$
  Actualizar los parámetros $\boldsymbol{w} \leftarrow \boldsymbol{w} - \alpha \nabla_{\boldsymbol{w}} J(\boldsymbol{w})$
end

Regresión multivariada

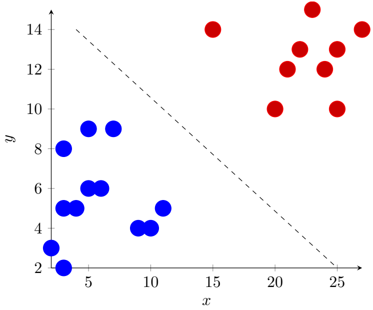


- La regresión lineal es aplicada a datos multivariados.
- En este caso, la predicción se calcula como combinación lineal de los atributos de entrada.

$$\hat{y} = w_0 + \sum_{j=1}^{m} w_j * x_j$$

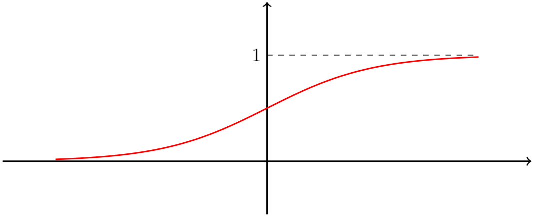
- Los parámetros del modelo se estiman a partir de los conceptos de Función de costo y gradiente descendente.

Regresión Logística (Clasificación binaria) I



- La regresión logística realiza predicciones a partir de una combinación lineal de los atributos.
- En la regresión logística se mide la probabilidad de que una instancia pertenezca a una de las dos clases (clasificación binaria).
- A diferencia de la regresión lineal, en la regresión logística (clasificación binaria), la idea es encontrar un hiper-plano que divida la región de entrada en dos (una por cada clase).

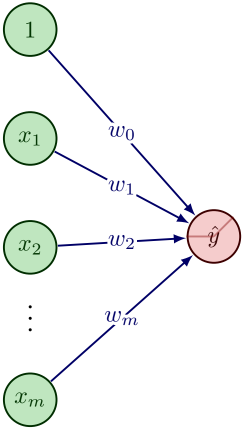
Regresión Logística (Clasificación binaria) III



- La función sigmoidal tiene como entrada los número reales y como salidas valores en el intervalo (0,1).
- Esto le permite tener a las predicciones un sentido de probabilidad.
- La función sigmoidal se representa a partir de la siguiente ecuación,

$$\sigma(a) = \frac{1}{1 + \exp(-a)}$$

Regresión Logística (Clasificación binaria) II



- Al igual que en la regresión lineal, la regresión logística se basa en una combinación lienal.
- La combinación lineal toma cualquier valor real. Así, para darle sentido de probabilidad a la predicción, se usa la función sigmoidal ($\sigma(\cdot)$).

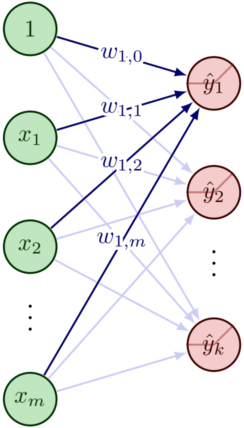
$$\hat{y} = p(y = 1|x) = \sigma\left(w_0 + \sum_{j=1}^{m} w_j * x_j\right)$$

Regresión Logística (Clasificación binaria) II

- La naturaleza de las etiquetas en un problema de clasificación difiere completamente de las tareas de regresión; así, la función de costo debe adaptarse a este tipo de etiquetas.
- Para la clasificación binaria, la función de costo más usada es la entropía cruzada, definida por,

$$\mathcal{L}(\mathbf{w}) = -\sum_{i=1}^{N} \left(y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right).$$

Regresión Logística (Clasificación mutlticlase)



- La regresión logística pueden extenderse para problemas de clasificación de múltiples clases.
- Se deben definir K salidas, cada salida está asociada con una de las clases.
- Con el fin de darle sentido de probabilidad a las predicciones, se usa la función softmax, la cual transforma la salida de la combinación lineal a valores entre 0 y 1.

$$\hat{y}_k = p(y = k|x), \quad \sum_{k=1}^{K} \hat{y}_k = 1.$$

- Ver el siguiente **tutorial**

Actividad

- Construir, paso a paso, un modelo de regresión logística.
- Probar el modelo construido sobre la base de datos `adult`