SGDClassifier

Problema de optimización

Dado un conjunto de entrenamiento $\{(x_i, y_i)\}_{i=1}^{n}$ donde $x_i \in \mathbb{R}^d$ y $y_i \in \{-1, 1\}$:

$$\min_{w,b} \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}(y_i, w^T x_i + b) + \alpha R(w)$$

Donde:

- $\mathcal{L}(y, \hat{y}) \longrightarrow$ Función de pérdida convexa
- $R(w) \longrightarrow$ Término de regularización como:
  - L2: $\frac{1}{2}\|w\|_2^2$
  - L1: $\|w\|_1$
  - Elastic Net: L1 y L2
- $\alpha > 0 \longrightarrow$ Parámetro de regularización

Modelo

SGDClassifier entrena un modelo lineal de la forma:

$$f(x) = \text{Sign}(w^T x + b)$$

$W \to$ Vector de pesos entrenados
$b \to$ Sesgo

La predicción se realiza con base en el signo del margen lineal.

Logistic Regression

Modelo

Dado un conjunto de entrenamiento $\{(x_i, y_i)\}_{i=1}^{n}$ con:

- $x_i \in \mathbb{R}^d$
- $y_i \in \{0,1\}$

$$P(y_i = 1 \mid x_i) = \sigma(W^T x_i + b)$$

donde $\sigma(z)$ es la función sigmoide:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

La predicción se hace con:

$$\hat{y}_i = \begin{cases} 1 & \text{Si } \sigma(W^T x_i + b) \geq 0.5 \\ 0 & \text{en otros casos} \end{cases}$$

Problema de optimización

$$\min_{w,b} \frac{1}{N} \sum_{i=1}^{n} \left[-y_i \log(\hat{p}_i) - (1-y_i) \log(1-\hat{p}_i) \right] + \alpha R(w)$$

donde:

$$\hat{p}_i = \sigma(w^T x_i + b)$$

$R(w) \rightarrow$ Regularización

Linear discrimination Analysis

Se asume que cada clase $y \in \{1, 2, \dots, K\}$ tiene una distribución gaussiana con:

$$x \mid y = k \sim \mathcal{N}(\mu_k, \Sigma)$$

Se asume:

- Covarianzas iguales $\Sigma_1 = \cdots = \Sigma_k = \Sigma$
- Vectores de medias por clase μ_k
- Priori de clase $\pi_k = P(y = k)$

La regla de decisión se basa en la función discriminante lineal:

$$\delta(x) = x^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \log \pi_k$$

Se asigna la clase:

$$\hat{y} = \arg\max_k \delta_k(x)$$

LDA proyecta los datos en un espacio lineal donde la separación entre clases es óptima según la razón de fisher:

$$J(w) = \frac{w^T S_B w}{w^T S_W w}$$

donde:

S_B = Matriz de dispersión entre clases

S_W = Matriz de dispersión dentro de clases

Problema de optimización:

Con reducción de dimensionalidad (2 clases):

$$\max_w \frac{(w^T(\mu_1 - \mu_2))^2}{w^T S_W w}$$

Para multiclase se proyectan los datos sobre un sub-espacio de dimensión $\leq K-1$ y se encuentran los vectores propios más relevantes de $S_W^{-1} S_B$

K Neighbors Classifier

Dado un conjunto de datos $\{(x_i, y_i)\}_{i=1}^{n}$

- $x_i \in \mathbb{R}^d$
- $y_i \in \{1, 2, \dots, K\}$

$$\hat{y} = \text{modo}(\{y_j \in N_K(x)\})$$

donde

- $N(x) \to$ Conjunto de los K vecinos más cercanos a x según una métrica de distancia
- $\text{modo}() \to$ Clase más frecuente entre esos vecinos

Distancia típica

Para 2 puntos x y z, la distancia euclidea es:

$$d(x, z) = \|x - z\|_2 = \sqrt{\sum_{i=1}^{d} (x_i - z_i)^2}$$