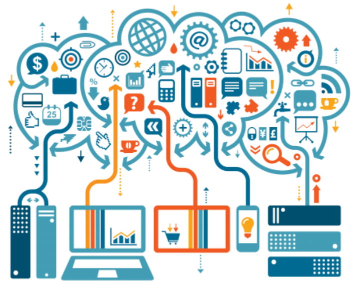


Minería de datos

Ing. Luis Francisco López

ESCUELA COLOMBIANA DE INGENIERÍA
JULIO GARAVITO

Clasificación (III)

Regresión logística

Logistic Regression

Concepto

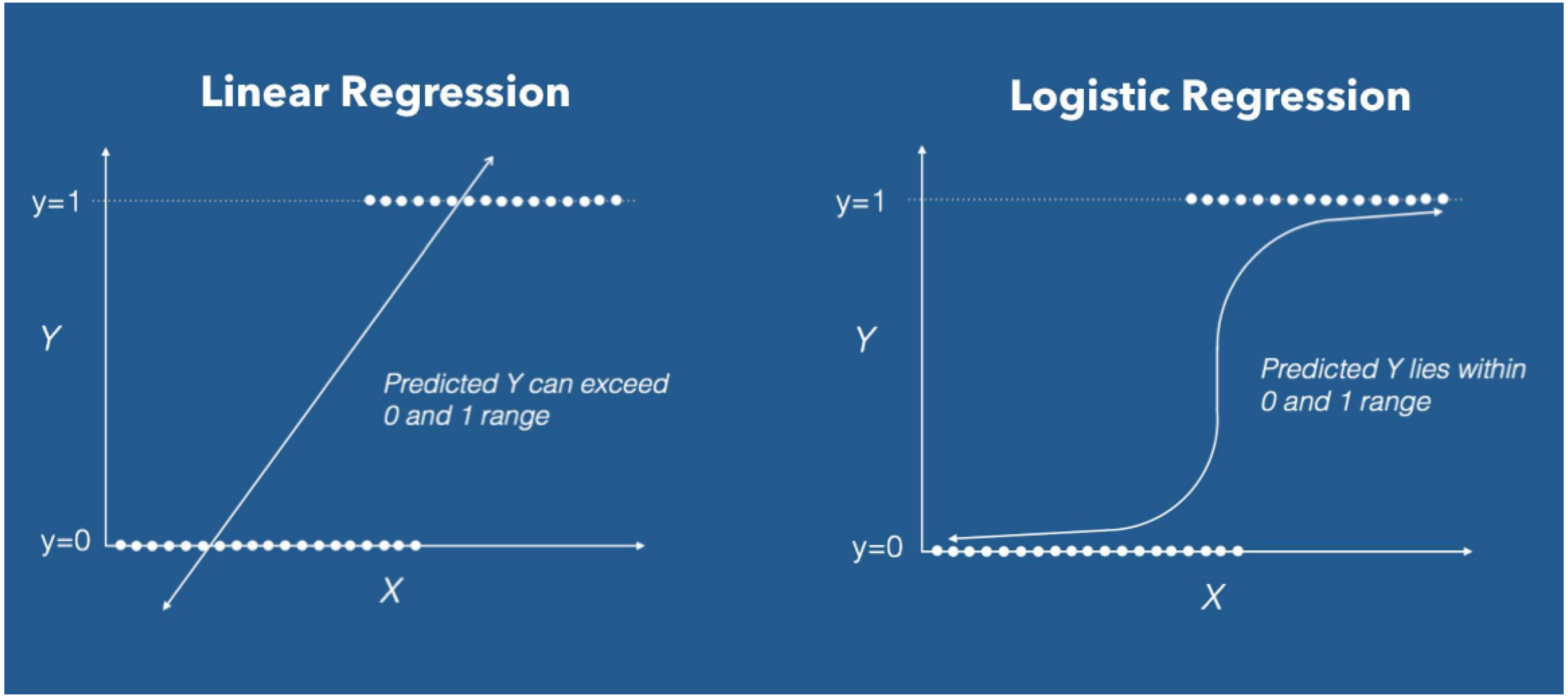
Supongamos que $P(X) = P(Y = 1|X)$ es la probabilidad de pertenecer a la clase 1 (en una clasificación binaria) a partir de una variable explicativa X.

La regresión logística usa la forma:

$$P(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 - e^{\beta_0 + \beta_1 X}}$$

Sin importar los valores de β_0, β_1 o X, $P(X)$ tendrá valores entre 0 y 1.

Regresión lineal vs. Regresión logística



Fuente: https://blog.knoldus.com/machinex-simplifying-logistic-regression/amp/

Concepto

Transformando:

Función lineal

$$\ln\left(\frac{P(X)}{1-P(X)}\right) = \beta_0 + \beta_1 X$$

Odds

Esta transformación monótona de se llama *Transformación log-odds* o *logit*.

Probabilidades y odds

Los "Odds" son usados en apuestas, por ejemplo, una apuesta 10 a 1 en favor de un equipo, significa que el corredor de la apuesta considera que la probabilidad de que dicho equipo no ganará es 10 veces la probabilidad de que si lo hará.

En términos de probabilidad:

$$\mathrm{Odds}\left(A\right) = \frac{P\left(A\right)}{1 - P\left(A\right)}$$

Regresión logística múltiple

$$P(X) = \frac{e^{\beta_0+\beta_1 X_1+\cdots+\beta_p X_p}}{1 - e^{\beta_0+\beta_1 X_1+\cdots+\beta_p X_p}}$$

$$\ln\left(\frac{P(X)}{1-P(X)}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

Regresión logística multiclase

La regresión logística puede generalizarse para más de dos clases:

$$P\left(Y=k|X\right)=\frac{e^{\beta_{0k}+\beta_{1k}X_1+\cdots+\beta_{pk}X_p}}{\sum_{l=1}^{K}e^{\beta_{0l}+\beta_{1l}X_1+\cdots+\beta_{pl}X_p}}$$

En este caso hay una función lineal por cada clase.

Son necesarias solamente $K-1$ funciones lineales.

Estimación de coeficientes

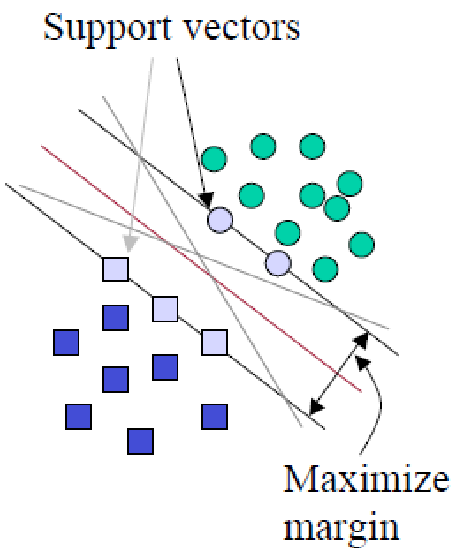
- Mínimos cuadrados (transformación logit)
- Estimación por máxima verosimilitud (MLE)
- Descenso más empinado (Steepest descent)
- Iteratively Reweighted Least Squares (IRLS)
- Stochastic average gradient (SAG)

Máquinas de soporte vectorial

Support Vector Machines (SVM)

Vectores de soporte

- Son los puntos de datos más cercanos a la superficie de decisión o hiperplano.
- Son los puntos de datos más difíciles de clasificar.
- Tienen relación directa en la localización dela superficie de decisión.



Fuente: An Idiot's guide to Support vector machines (SVMs). R. Berwick. MIT

Máquina de soporte vectorial (SVM)

- Maximiza el margen alrededor del hiperplano separador o superficie de decisión.
- La función de decisión está completamente especificada por un subconjunto (usualmente muy pequeño) de datos de entrenamiento llamados vectores de soporte.
- Es un problema de programación cuadrática que se resuelve por métodos estándar.

Para el caso no lineal

La idea es trasladar los datos a un espacio de mayor dimensionalidad donde le hiperplano es lineal y por lo tanto más fácil de calcular.

Transformaciones o Kernels:

- Polinomial

$$K\left(\boldsymbol{x}, \boldsymbol{y}\right) = \left(\mathbf{x} \cdot \mathbf{y} + 1\right)^p$$

- De base radial o Gaussiano

$$K\left(\boldsymbol{x}, \boldsymbol{y}\right) = \exp\left\{\frac{-\left\|\mathbf{x} - \mathbf{y}\right\|^2}{2\sigma^2}\right\}$$

- Tangente hiperbólica o sigmoide

$$K\left(\boldsymbol{x}, \boldsymbol{y}\right) = \tanh\left(k\mathbf{x} \cdot \mathbf{y} - \delta\right)$$