

# Tema 7

## Técnicas de clasificación

### Técnicas multivariantes

Dr. Antoni Ferragut



- La clasificación ajusta variables cualitativas.
- La regresión logística. Basado en modelo
- El método de los  $K$  vecinos más cercanos (KNN). Basado en ejemplos

- Para conseguir obtener valores de probabilidad de la variable respuesta usamos la función logística:

$$p(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}.$$

- Operando:

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

- El miembro de la izquierda son los **odds**, con valores en  $[0, \infty)$ .
- odds cerca de 0 (altos) corresponden a baja (alta) probabilidad del suceso.
- Tomando logaritmos:

$$\log \left( \frac{p(X)}{1 - p(X)} \right) = \beta_0 + \beta_1 X.$$

- El miembro de la izquierda son ahora los **log-odds** o **logit**.

## Estimación de los coeficientes:

- Estimamos  $\beta_0, \beta_1$  con datos de entrenamiento.
- Usamos el método de máxima verosimilitud  $\ell(\beta_0, \beta_1)$ : la probabilidad predicha  $\hat{p}(x_i)$  para cada individuo se acerca lo máximo posible a su valor observado.
- Escogemos la estimación de los coeficientes  $\hat{\beta}_0, \hat{\beta}_1$  que maximicen la ecuación:

$$\ell(\beta_0, \beta_1) = \prod_{i:y_i=1} p(x_i) \prod_{i:y'_i=0} (1 - p(x'_i)).$$

- El método de máxima verosimilitud es muy general y se usa en gran cantidad de modelos no lineales.
- Una vez estimados los coeficientes podemos realizar predicciones a partir de la expresión de  $p(X)$ .

- Podemos generalizar las ecuaciones anteriores:

$$p(X) = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}.$$

$$\log \left( \frac{p(X)}{1 - p(X)} \right) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p.$$

- El KNN se basa en las propias observaciones.
- Clasifica las nuevas observaciones basándose en las observaciones ya existentes (de entrenamiento).
- Se identifican los  $K > 0$  puntos más cercanos a una observación  $x_0$  que se quiere clasificar.
- Sea  $N_0$  este conjunto de puntos. Entonces:

$$P(Y = j|X = x_0) = \frac{1}{K} \sum_{i \in N_0} I(y_i = j).$$

La función  $I$  cuenta los valores de  $y_i$  que toman el valor  $j$ .

- Se asigna a  $x_0$  la clase que ha obtenido mayor probabilidad.
- Con  $K$  pequeño las fronteras entre regiones son más flexibles.
- Podemos usar validación cruzada para obtener un  $K$  óptimo que minimice el error de clasificación.

### **Pasos a realizar:**

- Estandarizar las variables predictoras: necesitamos una misma escala para comparar variables (distancias).
- Ajustar  $K$  de manera que minimice el error de clasificación (validación cruzada).
- Construir el clasificador con el valor de  $K$  óptimo para poder predecir nuevas observaciones.