Apoyo Ayudantía
Métricas de Desempeño

Regresión Logística

La función de enlace que nos permite relacionar la esperanza de la distribución Bernoulli con los predictores es:

$$logit(\pi_{\boldsymbol{x}_i}) = \ln\left(\frac{\pi_{\boldsymbol{x}_i}}{1-\pi_{\boldsymbol{x}_i}}\right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$$

Función de enlace → predictor lineal

Si en lo anterior aplicamos la función exponencial en ambos lados tenemos que:

$$\left(\frac{\pi_{\boldsymbol{x}_i}}{1-\pi_{\boldsymbol{x}_i}}\right) = \exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k) \quad (1)$$

El término $\frac{\pi_{\boldsymbol{x}_i}}{1-\pi_{\boldsymbol{x}_i}}$ se conoce como chance (odds). Las chances representan la relación entre la probabilidad de ocurrencia y no ocurrencia de un éxito.

- Si las chances con mayores a 1 la probabilidad de ocurrencia del éxito es mayor a la probabilidad de fracaso (más probable que ocurra el éxito).
- Si las chances son menores a 1 la probabilidad de ocurrencia del éxito es menor a la probabilidad del fracaso (menos probable que ocurra el éxito).

Regresión Logística

De la misma manera:

$$P(\widehat{y_i = 1}|x_i) = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k))}$$

Luego, si tenemos estimadores $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ podremos calcular $P(\widehat{y_i = 1}|x_i)$. Con esto podemos hacer una predicción:

$$\hat{Y} = \begin{cases} 1 & \text{si } P(\widehat{y_i = 1}|x_i) \geq c, \\ 0 & \text{si } P(\widehat{y_i = 1}|x_i) < c. \end{cases}$$

En esto c representa un punto de corte elegido por el usuario en función del problema. El punto de corte es un valor entre 0 y 1, tal que, si la probabilidad de un sujeto es mayor que este, se clasificará como positivo y en caso contrario como negativo.

El punto de corte dependerá del área de trabajo, ya que, por ejemplo, al disminuir el punto de corte, serán más los casos positivos detectados.

Regresión Logística

La razón entre dos chances se conoce como Odds Ratio (OR), o simplemente razón de chances. Esta cuantifica cuánto más probable es la ocurrencia de un evento al aumentar en una unidad o categoría una variable X_i específica.

Sean $\boldsymbol{x}_\star = (1, x_1, \ldots, x_i + 1, \ldots, x_k)$ y $\boldsymbol{x}_\bullet = (1, x_1, \ldots, x_i, \ldots, x_k)$, entonces la razón de chances está dada por:

$$OR = \frac{\left(\frac{\pi_{\boldsymbol{x}_\star}}{1-\pi_{\boldsymbol{x}_\star}}\right)}{\left(\frac{\pi_{\boldsymbol{x}_\bullet}}{1-\pi_{\boldsymbol{x}_\bullet}}\right)} = \frac{\exp(\beta_0 + \beta_1 x_1 + \ldots + \boldsymbol{\beta_i(x_i + 1)} + \ldots + \beta_k x_k)}{\exp(\beta_0 + \beta_1 x_1 + \ldots + \boldsymbol{\beta_i x_i} + \ldots + \beta_k x_k)} = \exp(\beta_i)$$

Con esto los coeficientes se interpretan de la siguiente forma:

- Si $\beta_i > 0$, entonces $OR > 1$ y por tanto X_i es un factor de riesgo.
- Si $\beta_i = 0$, entonces $OR = 1$ y por tanto no hay efecto del factor X_i.
- Si $\beta_i < 0$, entonces $OR < 1$ y por tanto X_i es un factor protector.

Matriz de confusión

La matriz de confusión (tambien conocidas como tablas de clasificación) (Matriz de Confusión) entregan información para evaluar la capacidad predictiva del modelo, hay que tener en consideración lo siguiente:

1. Un modelo puede ser correcto y tener malas propiedades de clasificación
2. En general, modelos con probabilidades estimadas cercanas a 0.5 tendrán bajo poder de clasificación.
3. Capacidad predictiva depende del punto de corte.

En las matrices de confusión se cruzan las predicciones con las tablas reales en una tabla de 2×2. En lo siguiente positivo se refiere a los éxitos (1) y negativos se refiere a los fracasos (0):

		Valores Reales	
		Positivos	**Negativos**
Valores Predichos	**Positivos**	Verdaderos Positivos (VP)	Falsos Negativos (FN)
	Negativos	Falsos Positivos (VP)	Verdaderos Negativos (VN)

Desde una matriz de confusión se pueden extraer diversos indicadores que nos permiten evaluar el modelo:

- **Sensibilidad:** = $\frac{VP}{VP+FN}$
- **Especificidad:** = $\frac{VN}{VN+FP}$
- **Precisión** = $\frac{VP}{VP+FP}$
- **Exactitud** = $\frac{VP+VN}{VP+VN+FP+FN}$

La **sensibilidad** representa la proporción de positivos capturados correctamente por el modelo, sobre el total de positivos reales. Representa que tan bien el modelo califica los casos "positivos" de nuestros datos.

La **especificidad** representa la proporción de casos negativos capturados correctamente por el modelo, sobre el total de negativos reales. Representa que tan bien el modelo califica los casos "negativos" de nuestros datos.

La **exactitud** mide la proporción de casos clasificados correctamente, independiente de si es positivo o negativo. Esta medida no es recomendada cuando la base de datos es desbalanceada.

La **presición** es el porcentaje de casos positivos clasificados correctamente.

Curva ROC

Una curva ROC representa la tasa de verdaderos positivos (sensibilidad) frente a la tasa de falsos positivos (1-especificidad) en diferentes umbrales de clasificación. Reducir el umbral de clasificación clasifica más elementos como positivos, por lo que aumentarán tanto los falsos positivos como los verdaderos positivos.

Desde la curva ROC se obtiene el indicador AUC (área bajo la curva), este representa el área bajo la curva ROC, entregando valores entre 0.5 (50%) y 1 (100%). Una forma de interpretar el AUC es la probabilidad que el modelo asigne una probabilidad más alta a un caso positivo que un caso negativo.