Los Modelos de Regresión y de Clasificación

Materia: Simulación de procesos

Prof. D.Sc. BARSEKH-ONJI Aboud

Facultad de Ingeniería

Universidad Anáhuac México

28 de enero de 2026

Agenda

¿Qué es la Regresión?

Objetivo

La regresión se enfoca en aprender la relación entre un conjunto de variables de entrada (predictoras) $\mathbf{x}$ y una (o varias) variable(s) de salida **cuantitativa** (respuesta) y.

¿Qué es la Regresión?

El objetivo es construir un modelo f que describa esta relación, generalmente expresado como:

$$y = f(\mathbf{x}) + \epsilon$$

Donde ϵ representa el término de error o ruido que no puede ser explicado por $\mathbf{x}$.

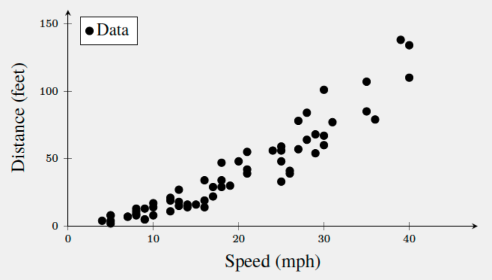


Figura 1: Ejemplo: Predecir la distancia de frenado a partir de la velocidad.

Regresión Lineal: El Modelo

Definición

Describe la variable de salida y como una combinación afín de las variables de entrada más un término de ruido. Es uno de los enfoques más fundamentales y utilizados.

Ecuación del Modelo

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon$$

- β_0 es el término de intercepto (sesgo).
- $\beta_1, \ldots, \beta_p$ son los coeficientes de regresión (pesos) que el modelo debe aprender a partir de los datos.

Propósitos de la Regresión Lineal

Describir Relaciones (Inferencia)

Interpretar los coeficientes β_j para entender la relación entre cada entrada x_j y la salida y, realizando pruebas de hipótesis para ver si la relación es estadísticamente significativa.

Predecir Salidas (Aprendizaje Automático)

Una vez aprendidos los coeficientes $\hat{\beta}$, se puede predecir una nueva salida $\hat{y}_*$ para una nueva entrada $\mathbf{x}_*$:

$$\hat{y}_* = \hat{\beta}_0 + \hat{\beta}_1 x_{*1} + \cdots + \hat{\beta}_p x_{*p}$$

Propósitos de la Regresión Lineal

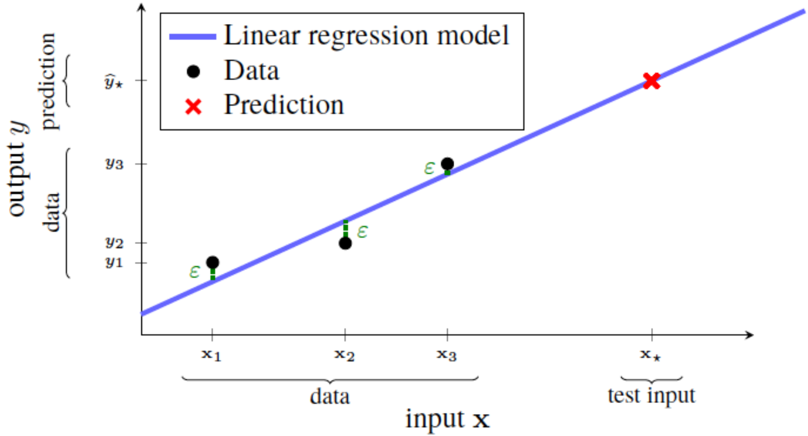


Figura 2: Modelo de regresión lineal simple ($p = 1$).

El Proceso de Aprendizaje

Consiste en estimar los parámetros β a partir de un conjunto de datos de entrenamiento. En forma matricial, el modelo se expresa como:

$$Y = X\beta + \epsilon$$

Aprendizaje: Máxima Verosimilitud y Mínimos Cuadrados

Mínimos Cuadrados (Least Squares)

Bajo el supuesto de que los errores siguen una distribución Gaussiana, el método de **Máxima Verosimilitud** es equivalente a minimizar la Suma de los Errores al Cuadrado (SSE). Este es el famoso criterio de **Mínimos Cuadrados**:

$$\hat{\beta}_{LS} = \arg\min_{\beta} ||Y - X\beta||_2^2$$

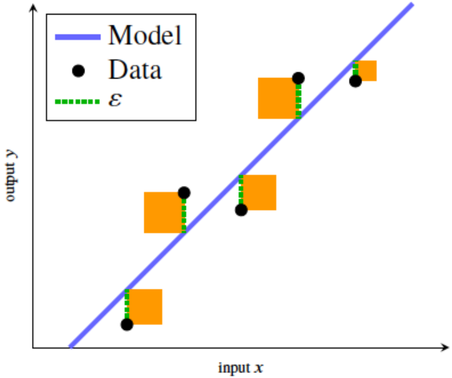


Figura 3: El criterio de Mínimos Cuadrados minimiza la suma de las áreas de los cuadrados naranjas.

La Solución: Ecuaciones Normales

Obtención de los Coeficientes

La solución al problema de mínimos cuadrados se obtiene al establecer el gradiente de la Suma de Errores al Cuadrado (SSE) a cero. Esto conduce a un sistema de ecuaciones lineales conocido como las **Ecuaciones Normales**.

$$X^T X \hat{\beta} = X^T Y$$

Solución de Forma Cerrada

Si la matriz $X^T X$ es invertible, existe una solución única de forma cerrada para los coeficientes $\hat{\beta}$:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

¿Qué Significa 'Lineal' en Regresión Lineal?

Linealidad en los Parámetros β

El término 'lineal' se refiere a que el modelo es una combinación lineal de los **parámetros** (β), no necesariamente de las variables de entrada originales.

Modelando Relaciones Complejas

Podemos incluir transformaciones no lineales de las entradas (creando nuevas características o *features*) para modelar relaciones más complejas. Por ejemplo, con una regresión polinomial:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_d x^d + \epsilon$$

Este sigue siendo un modelo de regresión lineal porque es lineal en los coeficientes β_j.

Ejemplo: Regresión Polinomial

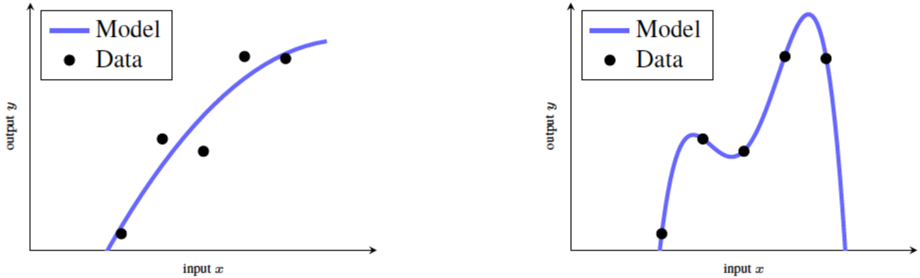


Figura 4: (izquierda) Modelo con un polinomio de segundo orden. (derecha) Modelo con un polinomio de cuarto orden. Un orden muy alto puede llevar a un sobreajuste.

Otras Transformaciones: Funciones de Base Radial (RBFs)

Funciones de Base Radial (RBFs)

Son otra clase popular de transformaciones, como el kernel Gaussiano:

$$K_c(x) = \exp\left(-\frac{||x - c||_2^2}{l}\right)$$

El modelo se convierte en una combinación lineal de estas funciones de base:

$$y = \beta_0 + \sum_{j=1}^{p} \beta_j K_{c_j}(x) + \epsilon$$

Propiedad Clave de las RBFs

Tienen un efecto **local**, lo que significa que un cambio en un parámetro afecta principalmente al modelo en la vecindad del centro c de la RBF.

Manejo de Variables Cualitativas (Categóricas)

La Técnica: Variables Ficticias (Dummy Variables)

Para incorporar variables de entrada cualitativas en un modelo de regresión lineal, se transforman en un conjunto de variables numéricas binarias (0 o 1) llamadas variables *dummy*.

Regla General

Si una variable cualitativa tiene K niveles o clases, se crean $K-1$ variables *dummy*. El nivel que no tiene una variable *dummy* se convierte en la categoría de referencia y queda capturado por el intercepto (β_0) del modelo.

Manejo de Variables Cualitativas (Categóricas)

Examples

Ejemplo: 'Tipo de Motor' Si tenemos una variable `TipoMotor` con 3 niveles: $\{\texttt{A, B, C}\}$.

- Se elige `A` como nivel de referencia.
- Se crean 2 variables *dummy* ($K - 1 = 3 - 1 = 2$):
  - x_{TipoB}: será 1 si `TipoMotor=B`, y 0 en otro caso.
  - x_{TipoC}: será 1 si `TipoMotor=C`, y 0 en otro caso.

Regularización: Evitando el Sobreajuste

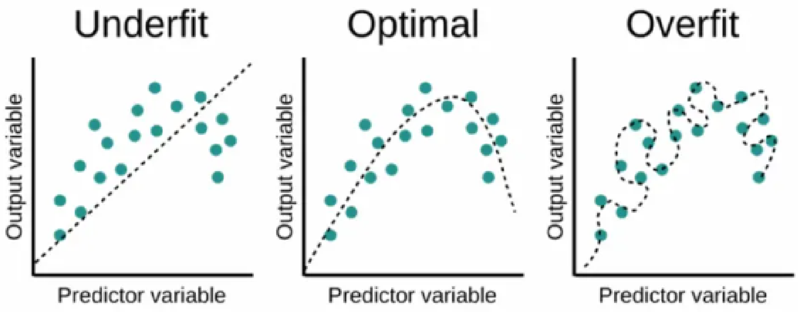


Figura 5: underfitting vs. overfitting

Regularización: Evitando el Sobreajuste

El Problema: Sobreajuste (Overfitting)

Cuando el número de características es grande comparado con el número de muestras, o cuando las características están muy correlacionadas, el modelo puede aprender el ruido"de los datos de entrenamiento. Esto se conoce como sobreajuste y resulta en un mal rendimiento con datos nuevos.

La Solución: Regularización

La regularización introduce un término de **penalización** en la función de costo para restringir la magnitud de los coeficientes β. Esto previene que el modelo se vuelva demasiado complejo y mejora su capacidad de generalización.

$$\min_{\beta} \underbrace{V(Y, X, \beta)}_{\text{Ajuste a los datos}} + \gamma \underbrace{R(\beta)}_{\text{Penalización}}$$

Técnicas de Regularización: Ridge vs. LASSO

Ridge Regression (L2)

- Añade una penalización proporcional a la suma de los **cuadrados** de los coeficientes.

$$\min_{\beta} \left(||X\beta - Y||_2^2 + \gamma ||\beta||_2^2 \right)$$

- **Efecto:** Encoge todos los coeficientes hacia cero, pero raramente los hace exactamente cero.
- Es muy útil cuando hay alta correlación entre predictores.

LASSO (L1)

- Añade una penalización proporcional a la suma de los **valores absolutos** de los coeficientes.

$$\min_{\beta} \left(||X\beta - Y||_2^2 + \gamma ||\beta||_1 \right)$$

- **Efecto:** Puede forzar a que algunos coeficientes sean **exactamente cero**.
- Realiza una selección automática de características, produciendo modelos más simples (dispersos).

Efecto Visual de la Regularización

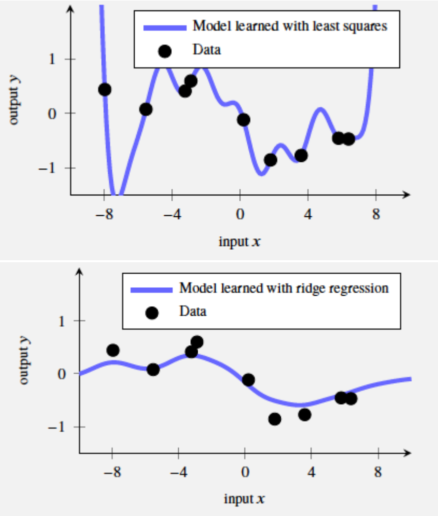


Figura 6: Comparación: (arriba) Sin regularización (sobreajuste), (abajo) Ridge (encoge coeficientes)

Efecto Visual de la Regularización

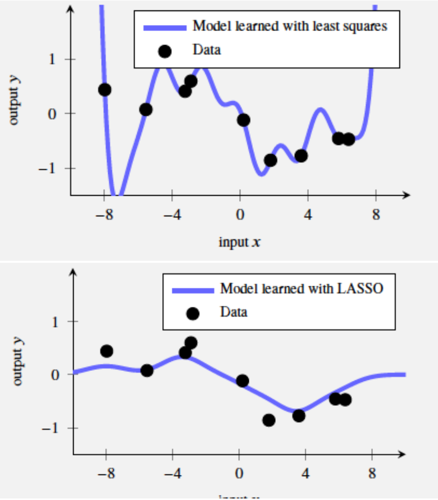


Figura 7: Comparación: (arriba) Sin regularización (sobreajuste), (abajo) LASSO (crea un modelo disperso).

Agenda

¿Qué es la Clasificación?

Objetivo

La clasificación es la tarea de predecir una variable de salida **cualitativa** (o categórica) y a partir de un conjunto de variables de entrada $\mathbf{x}$.

- La salida y puede tomar valores de un conjunto finito de clases o etiquetas: $\{c_1, c_2, \ldots, c_K\}$.
- **Ejemplos:** Clasificar correos como 'spam' o 'no spam' ($K = 2$); reconocer dígitos escritos a mano del 0 al 9 ($K = 10$).

Perspectiva Estadística

El objetivo principal es modelar o estimar las probabilidades de que una observación pertenezca a cada clase, dadas sus características:

$$p(y = c_k|\mathbf{x})$$

Regresión Logística: Un Clasificador Lineal

¿Qué es la Regresión Logística?

A pesar de su nombre, es un modelo de **clasificación lineal** muy popular. En lugar de modelar la salida y directamente, modela la **probabilidad** de que y pertenezca a una clase particular.

Modelo para Clasificación Binaria

Para $y \in \{0, 1\}$, modela la probabilidad de la clase positiva ($y = 1$) usando la función logística (o sigmoide):

$$P(y = 1|\mathbf{x}) = \frac{1}{1 + e^{-\mathbf{x}^T\beta}}$$

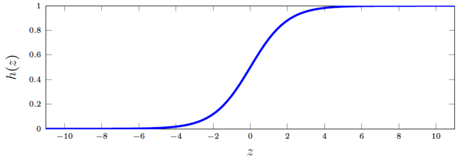


Figura 8: La función logística mapea cualquier valor real al intervalo (0,1).

Aprendizaje del Modelo de Regresión Logística

Máxima Verosimilitud (Maximum Likelihood Estimation - MLE)

Los parámetros β se estiman maximizando la función de verosimilitud logarítmica, que mide qué tan bien los parámetros del modelo explican los datos de entrenamiento observados.

$$\ell(\beta) = \sum_{i=1}^{n} \left(y_i \mathbf{x}_i^T \beta - \log(1 + e^{\mathbf{x}_i^T \beta}) \right)$$

Solución Numérica

No existe una solución de forma cerrada para β que maximice esta función. Por lo tanto, se deben utilizar métodos de optimización numérica iterativos, como el algoritmo de Newton-Raphson, para encontrar los mejores coeficientes.

Fronteras de Decisión para Regresión Logística

Una vez entrenado el modelo, se asigna una clase predicha. Típicamente, si $P(y = 1|\mathbf{x}) > 0{,}5$, la predicción es 1.

La Frontera de Decisión

El umbral de decisión de 0.5 corresponde a la ecuación:

$$\mathbf{x}^T \hat{\beta} = 0$$

Esta ecuación define un hiperplano lineal en el espacio de características. Por esta razón, la regresión logística es un **clasificador lineal**.

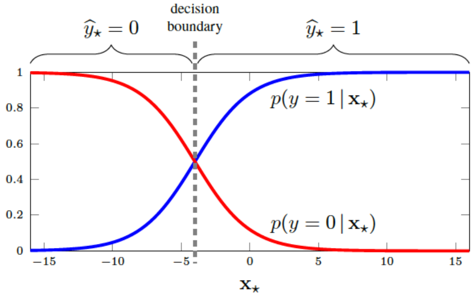


Figura 9: La frontera de decisión (línea vertical) separa las regiones donde la probabilidad predicha es ¿0.5 o ¡0.5.

Regresión Logística para Múltiples Clases (Multinomial)

Extensión a K $>$ 2 Clases

La regresión logística se puede generalizar a problemas multiclase mediante la **regresión softmax**. Se calcula la probabilidad para cada clase k usando la función softmax:

$$P(y = k|\mathbf{x}) = \frac{e^{\mathbf{x}^T \beta_k}}{\sum_{j=1}^{K} e^{\mathbf{x}^T \beta_j}}$$

Fronteras de Decisión Lineales

Incluso en el caso multiclase, las fronteras de decisión entre cualquier par de clases siguen siendo lineales.

Regresión Logística para Múltiples Clases (Multinomial)

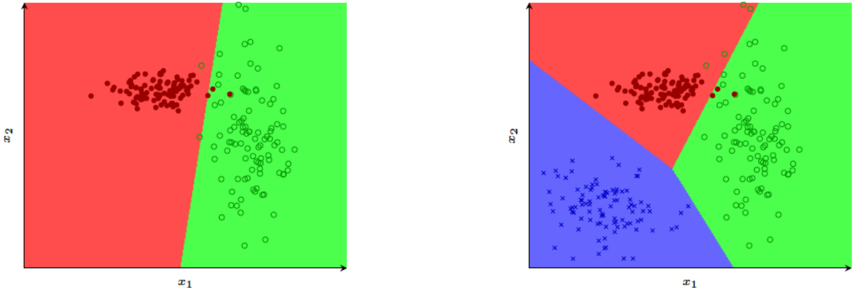


Figura 10: Fronteras de decisión para regresión logística con K=3 clases (izquierda) y K=2 clases con dos características (derecha).

LDA y QDA: Un Enfoque Generativo

Métodos de Clasificación Generativos

A diferencia de la regresión logística (que es un modelo *discriminativo*), LDA y QDA son modelos **generativos**. En lugar de modelar la frontera de decisión directamente, modelan la distribución de las características para cada clase.

El Teorema de Bayes en Acción

Modelan la probabilidad de las características dada la clase, $p(\mathbf{x}|y=k)$, y la probabilidad a priori de cada clase, $P(y=k)$. Luego, usan el Teorema de Bayes para calcular la probabilidad final que nos interesa:

$$P(y=k|\mathbf{x}) = \frac{p(\mathbf{x}|y=k)P(y=k)}{\sum_{j=1}^{K} p(\mathbf{x}|y=j)P(y=j)}$$

La Suposición Gaussiana: Diferencia entre LDA y QDA

Suposición Principal

Tanto LDA como QDA asumen que la distribución de las características para cada clase, $p(\mathbf{x}|y=k)$, sigue una distribución Gaussiana multivariada.

Análisis Discriminante Lineal (LDA)

- **Suposición Fuerte:** Todas las clases comparten la **misma** matriz de covarianza ($\Sigma_k = \Sigma$).
- **Resultado:** Genera fronteras de decisión **lineales**.

Examples

Análisis Discriminante Cuadrático (QDA)

- **Suposición Flexible:** Cada clase tiene su **propia** matriz de covarianza (Σ_k).
- **Resultado:** Genera fronteras de decisión **cuadráticas**.

La Suposición Gaussiana: Diferencia entre LDA y QDA

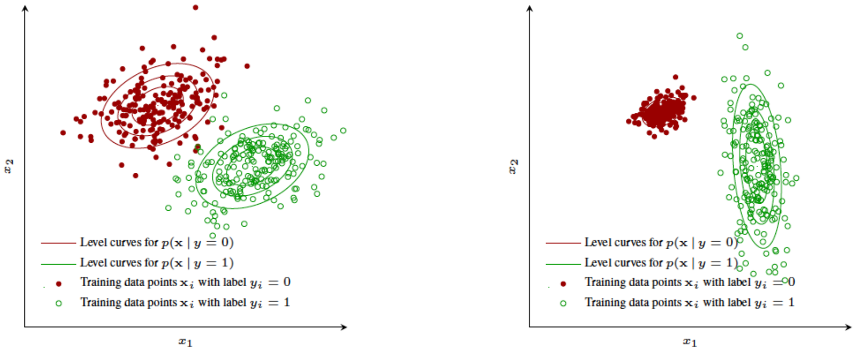


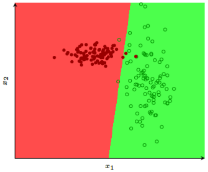
Figura 11: Diferencia en las suposiciones de covarianza para LDA (izquierda) y QDA (derecha).

Fronteras de Decisión Resultantes

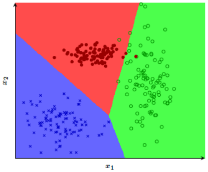
La Forma de la Frontera de Decisión

La suposición sobre la matriz de covarianza determina directamente la forma de la frontera que separa las clases.

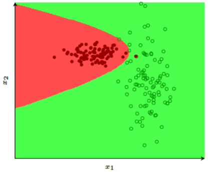
Fronteras de Decisión Resultantes



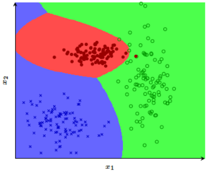
(a) LDA for $K = 2$ classes always gives a linear decision boundary. The red dots and green circles are training data from different classes, and the intersection between the red and green fields is the decision boundary obtained for an LDA classifier learned from the training data.



(b) LDA for $K = 3$ classes. We have now introduced training data from a third class, marked with blue crosses. The decision boundary between any two pair of classes is still linear.



(c) QDA has quadratic (i.e., nonlinear) decision boundaries, as in this example where a QDA classifier is learned from the shown training data.



(d) With $K = 3$ classes are the decision boundaries for QDA possibly more complex than with LDA, as in this case (cf. (b)).

Figura 12: Ejemplos de fronteras de decisión: (a, b) LDA siempre produce fronteras lineales. (c, d) QDA produce fronteras cuadráticas, que son más flexibles.

El Clasificador de Bayes: La Decisión Óptima

La Regla de Decisión Óptima

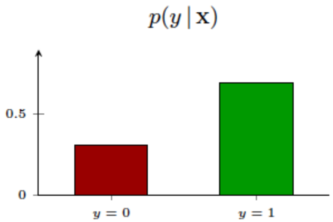
El Clasificador de Bayes proporciona la justificación teórica para la toma de decisiones en clasificación. Si el objetivo es minimizar el número promedio de errores, el clasificador óptimo es aquel que predice la clase $\hat{y}$ con la **mayor probabilidad a posteriori**.

$$\hat{y} = \arg\max_k P(y = k|\mathbf{x}_*)$$

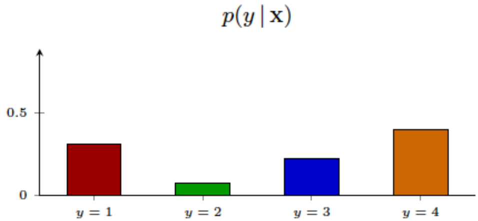
Consideraciones Prácticas

En la práctica, no conocemos las verdaderas probabilidades $P(y = k|\mathbf{x})$, por lo que los modelos de aprendizaje automático se utilizan para **estimarlas**. Además, esta regla asume que todos los errores cuestan lo mismo. Si los costos de error son asimétricos (p.ej., en diagnóstico médico), se deben incorporar funciones de pérdida.

El Clasificador de Bayes: La Decisión Óptima



(a) With $K = 2$ classes, Bayes' classifier tell us to take the class which has probability > 0.5 as the prediction $\widehat{y}$. Here, the prediction would therefore be $\widehat{y} = 1$.



(b) For $K = 4$ classes, Bayes' classifier tells us to take the prediction $\widehat{y}$ as the highest bar, which means $\widehat{y} = 4$ here. (In contrast to $K = 2$ classes in (a), it can happen that no class has probability > 0.5.)

Figura 13: El clasificador de Bayes selecciona la clase con la probabilidad más alta.

Consideraciones Adicionales

Clasificadores Lineales vs. No Lineales

- Un clasificador es **lineal** si su frontera de decisión es lineal (e.g., Regresión Logística, LDA).
- QDA es un clasificador **no lineal**.
- Se pueden obtener fronteras no lineales con clasificadores lineales mediante la transformación no lineal de las características de entrada.

Regularización en Clasificación

- Al igual que en la regresión, la regularización (L1 o L2) es crucial para evitar el sobreajuste en modelos de clasificación.
- Es especialmente útil cuando hay muchas características o los datos son limitados.

Más allá de la Exactitud (Accuracy)

El Problema con la Exactitud

La exactitud simple (porcentaje de aciertos) puede ser engañosa, especialmente si las clases están desbalanceadas (e.g., 99 % de los correos no son spam).

La Matriz de Confusión

Es la herramienta fundamental para resumir el rendimiento de un clasificador binario.

		Clase Real	
		Positivo	Negativo
Predicción	Positivo	**VP** (Verdadero Positivo)	**FP** (Falso Positivo)
	Negativo	**FN** (Falso Negativo)	**VN** (Verdadero Negativo)

Métricas Clave de la Matriz de Confusión

Sensibilidad (Recall)

¿Qué proporción de los positivos reales se identificó correctamente?

$$\text{Sensibilidad} = \frac{VP}{VP + FN}$$

Especificidad

¿Qué proporción de los negativos reales se identificó correctamente?

$$\text{Especificidad} = \frac{VN}{VN + FP}$$

Precisión (Precision)

De todos los que se predijeron como positivos, ¿cuántos lo eran realmente?

$$\text{Precisión} = \frac{VP}{VP + FP}$$

Métricas Clave de la Matriz de Confusión

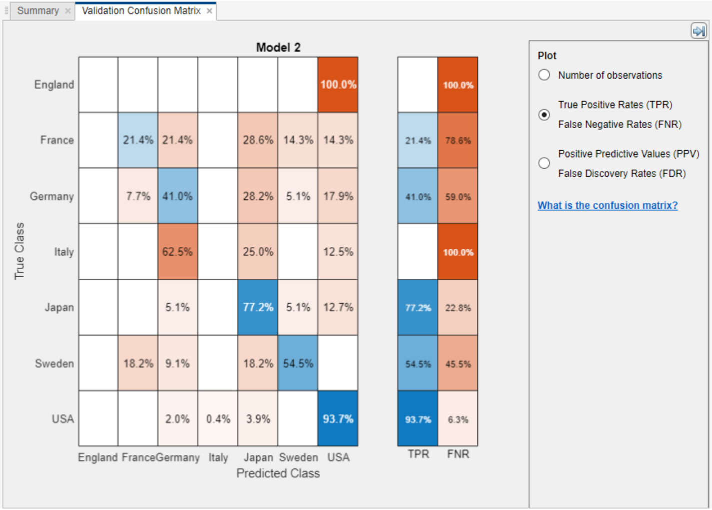


Figura 14: Ejemplo de una matriz de confusión

Curva ROC y AUC

Curva ROC (Receiver Operating Characteristic)

Es un gráfico que muestra el rendimiento de un clasificador para todos los umbrales de decisión. Grafica:

- Eje Y: **Sensibilidad** (Tasa de Verdaderos Positivos).
- Eje X: **1 - Especificidad** (Tasa de Falsos Positivos).

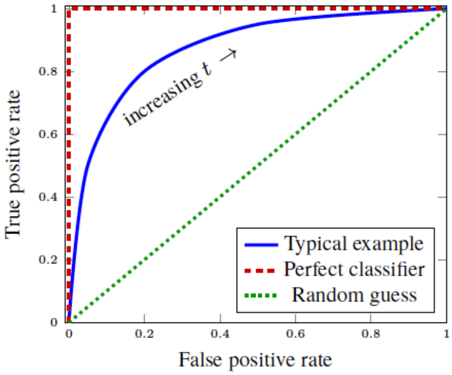


Figura 15: Ejemplo de una curva ROC.

Curva ROC y AUC

Área Bajo la Curva (AUC)

Es una medida global del rendimiento del clasificador, independiente del umbral.

- **AUC = 1:** Clasificador perfecto.
- **AUC = 0.5:** Clasificador aleatorio (inútil).

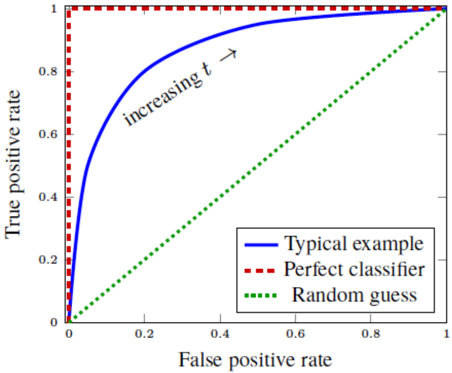


Figura 16: Ejemplo de una curva ROC.

Agenda

Ejemplos prácticos

Abordaremos los siguientes ejemplos:

- Pronóstico del tipo de cambio (USD/MXN) - Regresión 1
- Describir la relación entre varaibles que influyen en el rendimiento de un auto - Regresión 2
- Clasificación 'iris de Fisher 1936'.

Ejemplos prácticos

Para más información sobre el uso de las aplicaciones 'Regression Learner' y 'Classification Learner' de MATLAB es recomendable visitar:

- Para el caso de 'Regression Learner':
  https://la.mathworks.com/help/stats/regressionlearner-app.html
- Para el caso de 'Classification Learner':
  https://la.mathworks.com/help/stats/classificationlearner-app.html

Ejemplos prácticos

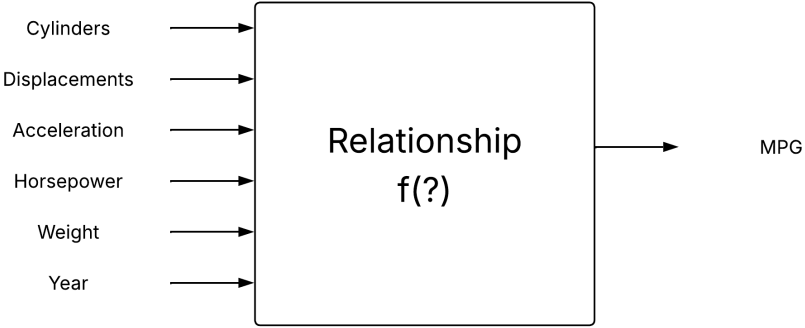


Figura 17: Modelo de regresión para el ejemplo 2

Ejemplos prácticos

Ejemplo de un modelo de clasificación

Este ejemplo utiliza los datos de iris de Fisher de 1936. Estos datos contienen mediciones de flores: longitud de pétalo, ancho de pétalo, longitud de sépalo y ancho de sépalo para especímenes de tres especies. Entrene un clasificador para predecir la especie basándose en las mediciones de los predictores.

Examples

Se recomienda revisar el ejemplo a detalle en el repositorio de Matlab en:
`https://la.mathworks.com/help/stats/train-decision-trees-in-classification-learner-app.html`