



UNIVERSITAS
Miguel Hernández

Tema 1. Regresión Lineal. Práctica.

José L. Sainz-Pardo Auñón

TÉCNICAS ESTADÍSTICAS PARA EL APRENDIZAJE II

Máster Universitario en Estadística Computacional
y Ciencia de Datos para la Toma de Decisiones.

Objetivo de la Práctica

- Utilizar diferentes técnicas de regresión para obtener el mejor modelo predictivo.
- Evaluar y comparar los modelos de regresión usando métricas estándar.
- Basar el análisis en bases de datos abiertas y reales para un enfoque práctico.

Base de Datos Utilizada

- Boston Housing Dataset (Disponible en 'scikit-learn'):
 - ▶ Predecir el precio de las viviendas en función de características socioeconómicas y geográficas.
- Bases de datos real que se usa frecuentemente para aplicar métodos de regresión.

1. Preprocesamiento de Datos

- Cargar los datos y manejar valores nulos, si es necesario.
- Realizar una exploración inicial:
 - ▶ Descripción estadística de las variables.
 - ▶ Gráficos de dispersión para evaluar relaciones entre las variables.

2. Selección de Variables

- Utilizar la matriz de correlación para identificar multicolinealidad.
- Aplicar el Factor de Inflación de la Varianza (VIF):
 - ▶ Eliminar variables altamente correlacionadas que aumenten la multicolinealidad.

3. Regresión Lineal Básica

- Ajustar un modelo de regresión lineal a los datos (utiliza el 80% de los datos como muestra de entrenamiento y el resto como conjunto de prueba).
- Evaluar el modelo usando:
 - ▶ Coeficientes de la regresión.
 - ▶ R^2 : Explicación de la varianza.
- Visualización:
 - ▶ Gráfico de los valores reales vs valores predichos.

5. Evaluación y Validación de Modelos

- Comparar los diferentes modelos basándose en:
 - ▶ R^2 ajustado.
- Analizar los residuos gráficamente y validar sus supuestos.

7. Mejora del Modelo

- Experimentar con diferentes transformaciones de variables como transformación logarítmica, cuadrática o raíz cuadrada de algunas variables.
- Probar nuevas interacciones entre variables que puedan mejorar el modelo.
- Añadir polinomios de mayor grado para capturar la no linealidad.

8. Más bases de datos

Puedes seguir realizando prácticas con las siguientes bases de datos:

- Auto MPG Dataset
(<https://archive.ics.uci.edu/dataset/9/auto+mpg>)
- Wine Quality Dataset
(<https://archive.ics.uci.edu/dataset/186/wine+quality>)
- Energy Efficiency Dataset (<https://archive.ics.uci.edu/dataset/242/energy+efficiency>)
- Concrete Compressive Strength (<https://archive.ics.uci.edu/dataset/165/concrete+compressive+strength>)



UNIVERSITAS
Miguel Hernández

Tema 1. Regresión Lineal. Práctica.

José L. Sainz-Pardo Auñón

TÉCNICAS ESTADÍSTICAS PARA EL APRENDIZAJE II

Máster Universitario en Estadística Computacional
y Ciencia de Datos para la Toma de Decisiones.

Introducción a la Regresión Logística

- La regresión logística es un modelo estadístico usado para predecir la probabilidad de una clase binaria.
- Fórmula básica:

$$P(y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}}$$

- Ejemplos de aplicación:
 - ▶ Predicción de enfermedades (enfermo/no enfermo).
 - ▶ Clasificación de correos electrónicos (spam/no spam).
 - ▶ Diagnóstico de cáncer (benigno/maligno).

Descarga y Carga de los Datos

- Descarga el dataset **Breast Cancer Wisconsin (Diagnostic)** desde el UCI Machine Learning Repository: [https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)).
- Carga los datos en tu entorno de programación utilizando Python. Asegúrate de que todas las variables se carguen correctamente.
- Revisa las primeras filas del dataset y obtén un resumen estadístico básico para entender la distribución de las características.

Explora los Datos

- Analiza la variable objetivo (Diagnosis) y asegúrate de identificar la proporción de casos benignos y malignos.
- Visualiza las distribuciones de las variables más importantes como Radius Mean, Texture Mean, y otras características clave.
- Convierte la variable Diagnosis a numérica: asigna 0 a los valores "B" (benigno) y 1 a los valores "M" (maligno).
- Dibuja un mapa de correlaciones entre las variables para identificar relaciones importantes.

Preprocesa los Datos

- Elimina la columna ID, ya que no aporta información relevante al modelo.
- Divide el dataset en conjunto de entrenamiento (70%) y de prueba (30%) utilizando la función `train_test_split`.

Ajusta el Modelo de Regresión Logística

- Entrena un modelo de regresión logística utilizando los datos del conjunto de entrenamiento.
- Analiza los coeficientes obtenidos.

Evalúa el Modelo

- Genera predicciones tanto para el conjunto de prueba como para el de entrenamiento utilizando el modelo ajustado.
- Visualiza la matriz de confusión (tanto en el conjunto de prueba como en el de entrenamiento) para analizar el rendimiento del modelo en términos de verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos.
- En ambos conjuntos, calcula las métricas de rendimiento como la precisión (accuracy), sensibilidad (recall) y precisión (precision).
- Traza la curva ROC y calcula el área bajo la curva (AUC) para evaluar la capacidad del modelo para distinguir entre clases.
- Prueba a realizar predicciones con algún umbral distinto a 0.5.

Interpreta los Resultados

- Calcula la exactitud del modelo y analiza los resultados obtenidos.
- Interpreta la matriz de confusión y el reporte de clasificación para identificar qué tan bien el modelo predice los tumores benignos y malignos.
- Revisa el valor del AUC y discute la capacidad del modelo para distinguir entre casos positivos y negativos.
- Reflexiona sobre posibles mejoras al modelo, como ajustar hiperparámetros o probar otros algoritmos de clasificación.
- Si normalizas los datos numéricos, ¿obtienes las mismas estimaciones?

Practica con Otros Datasets

- Repite la práctica con otros datasets disponibles en el UCI Machine Learning Repository para reforzar los conceptos de regresión logística.
- Algunas sugerencias:
 - ▶ **Pima Indians Diabetes Database:** <https://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes>
 - ▶ **Heart Disease Dataset:** <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>
 - ▶ **Bank Marketing Dataset:** <https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>
 - ▶ **Adult (Census Income) Dataset:** <https://archive.ics.uci.edu/ml/datasets/Adult>
- Sigue el mismo flujo de trabajo: carga los datos, explora las variables, preprocesa el dataset, ajusta un modelo de regresión logística y evalúa los resultados. Si hubiera variables **independientes categóricas**, descártalas del modelo.



UNIVERSITAS
Miguel Hernández

Tema 3. kNN. Práctica.

José L. Sainz-Pardo Auñón

TÉCNICAS ESTADÍSTICAS PARA EL APRENDIZAJE II

Máster Universitario en Estadística Computacional
y Ciencia de Datos para la Toma de Decisiones.

Descarga y Carga de los Datos

- Descarga el dataset **Breast Cancer Wisconsin (Diagnostic)** desde el UCI Machine Learning Repository: [https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)).
- Carga los datos en tu entorno de programación utilizando Python. Asegúrate de que todas las variables se carguen correctamente.
- Revisa las primeras filas del dataset y obtén un resumen estadístico básico para entender la distribución de las características.

Explora los Datos

- Analiza la variable objetivo (Diagnosis) y asegúrate de identificar la proporción de casos benignos y malignos.
- Visualiza las distribuciones de las variables más importantes como Radius Mean, Texture Mean, y otras características clave.
- Convierte la variable Diagnosis a numérica: asigna 0 a los valores "B" (benigno) y 1 a los valores "M" (maligno).
- Dibuja un mapa de correlaciones entre las variables para identificar relaciones importantes.

Preprocesa los Datos

- Elimina la columna ID, ya que no aporta información relevante al modelo.
- Divide el dataset en conjunto de entrenamiento (70%) y de prueba (30%) utilizando la función `train_test_split`.

Ajusta el Modelo de kNN

- Obtén a partir del conjunto de entrenamiento el k entre 1 y 31 con el que que mayor AUC se obtiene para el modelo k -NN (puedes utilizar la función `GridSearchCV`).
- Obtén los pronósticos de la muestra de prueba con el mejor k que obtuviste.
- Obtén el informe de clasificaciónn del modelo, utilizando la librería `sklearn`.
- Realiza distintas pruebas con distintos scores (accuracy, precision, ...)

Validación cruzada

- Realiza una validación cruzada utilizando 4 pliegues (4-fold, o 25% de ítems en la muestra de reserva) para evaluar el modelo con el conjunto de datos de entrenamiento.
- Calcula y muestra la media de las puntuaciones AUC obtenidas durante la validación cruzada.
- Obtén las predicciones sobre el conjunto de prueba.
- Obtén el informe de clasificación del conjunto de prueba, utilizando la librería sklearn.
- Obtén la curva ROC y el área bajo la misma (AUC).

Selección de prototipos mediante CNN.

- Utiliza Condensed Nearest Neighbor (CNN) para seleccionar prototipos de la muestra de entrenamiento.
- Graba en un fichero Excel el conjunto de entrenamiento y en otro fichero Excel el de prototipos obtenido. Observa ambos. ¿En qué porcentaje se realizó la reducción?
- Obtén las predicciones y el informe de clasificación utilizando el conjunto de prototipos. ¿Mejoró la proporción de acierto utilizando prototipos? Compara ambos informes de clasificación (con y sin prototipos).

Selección de prototipos mediante ENN.

- Utiliza Edited Nearest Neighbor (ENN) para seleccionar prototipos de la muestra de entrenamiento.
- Graba en un fichero Excel el conjunto de entrenamiento y en otro fichero Excel el de prototipos obtenido. Observa ambos. ¿En qué porcentaje se realizó la reducción?
- Obtén las predicciones y el informe de clasificación utilizando el conjunto de prototipos. Compara todos los informes de clasificación (sin prototipos, con prototipos CNN y con prototipos ENN). ¿En cuál de todos la proporción de acierto resultó mejor?

Practica con Otros Datasets

- Repite la práctica con otros datasets disponibles en el UCI Machine Learning Repository y ya conocidos tanto de Regresión como de Clasificación. Por ejemplo:
 - ▶ **Pima Indians Diabetes Database:**
<https://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes>
 - ▶ **Heart Disease Dataset:**
<https://archive.ics.uci.edu/ml/datasets/Heart+Disease>
 - ▶ **Bank Marketing Dataset:**
<https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>
 - ▶ **Adult (Census Income) Dataset:**
<https://archive.ics.uci.edu/ml/datasets/Adult>
 - ▶ **Auto MPG Dataset:** <https://archive.ics.uci.edu/dataset/9/auto+mpg>
 - ▶ **Wine Quality Dataset:**
<https://archive.ics.uci.edu/dataset/186/wine+quality>
 - ▶ **Energy Efficiency Dataset:**
<https://archive.ics.uci.edu/dataset/242/energy+efficiency>
 - ▶ **Concrete Compressive Strength:**
<https://archive.ics.uci.edu/dataset/165/concrete+compressive+strength>
- Sigue el mismo flujo de trabajo: carga los datos, explora las variables, preprocesa el dataset, ajusta el modelo y evalúa los resultados sin y con validación cruzada. Prueba también a seleccionar prototipos.
- Compara los resultados con aquellos obtenidos con Regresión Lineal o Logística.

Reconocimiento de género musical

Roni Bandini creó un algoritmo para reconocer el género musical que su vecino escuchaba y cuando la clasificaba como 'Reaggaetone' interfería en su emisión:
[https://es.euronews.com/next/2024/04/09/](https://es.euronews.com/next/2024/04/09/reggaetonbe-gone-esta-maquina-casera-silencia-la-musica-de-los-vec)

[reggaetonbe-gone-esta-maquina-casera-silencia-la-musica-de-los-vec](#)

El Spotify Tracks Dataset (spotify.csv) contiene un conjunto de canciones con sus características de audio, como energía, tempo, etc. recopilado de Spotify mediante su API. Tiene 114.000 registros de canciones y su género. Son muchos registros para aplicar kNN, no obstante puedes practicar, por ejemplo, con 1000 registros tomados al azar. Por cierto, en Kaggle existe un hilo sobre este Dataset:

<https://www.kaggle.com/datasets/maharshipandya/-spotify-tracks-dataset?resource=download>

en el que encontrarás la descripción de cada variable y código sobre métodos para tratarlo.

Propuesta de Trabajo Fin de Máster (TFM)

Como has podido ver, el Spotify Dataset Tracks contiene cientos de miles de canciones con diversas características acústicas y no es posible la aplicación de técnicas de clasificación sobre toda la base de datos.

El objetivo de esta propuesta de TFM es desarrollar un modelo de clasificación eficiente en el que se desarrollen técnicas avanzadas de selección de prototipos para reducir el tamaño del conjunto de entrenamiento sin sacrificar la precisión, mejorando así el rendimiento del modelo de clasificación.

Aquellas personas interesadas en realizar este TFM, enviad email a:
`jlsainz@umh.es`



UNIVERSITAS
Miguel Hernández

Tema 5. SVM. Práctica.

José L. Sainz-Pardo Auñón

TÉCNICAS ESTADÍSTICAS PARA EL APRENDIZAJE II

Máster Universitario en Estadística Computacional
y Ciencia de Datos para la Toma de Decisiones.

Descarga y Carga de los Datos

- Descarga el dataset **Breast Cancer Wisconsin (Diagnostic)** desde el UCI Machine Learning Repository: [https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)).
- Carga los datos en tu entorno de programación utilizando Python. Asegúrate de que todas las variables se carguen correctamente.
- Revisa las primeras filas del dataset y obtén un resumen estadístico básico para entender la distribución de las características.

Explora los Datos

- Analiza la variable objetivo (Diagnosis) y asegúrate de identificar la proporción de casos benignos y malignos.
- Visualiza las distribuciones de las variables más importantes como Radius Mean, Texture Mean, y otras características clave.
- Convierte la variable Diagnosis a numérica: asigna 0 a los valores "B" (benigno) y 1 a los valores "M" (maligno).
- Dibuja un mapa de correlaciones entre las variables para identificar relaciones importantes.

Preprocesa los Datos

- Elimina la columna ID, ya que no aporta información relevante al modelo.
- Divide el dataset en conjunto de entrenamiento (70%) y de prueba (30%) utilizando la función `train_test_split`.

Ajusta el Modelo de SVM

- Obtén a partir del conjunto de entrenamiento el C que mejor funciona para el modelo SVM . Prueba, por ejemplo, con el siguiente vector de búsqueda (0.001, 0.01, 0.1, 1, 10, 100 y 1000).
- Obtén los pronósticos de la muestra de prueba con el mejor C que obtuviste.
- Obtén el informe de clasificación del modelo, utilizando la librería `sklearn`.
- Realiza distintas pruebas con distintos scores (accuracy, precision, ...)

Validación cruzada

- Realiza una validación cruzada utilizando 4 pliegues (4-fold, o 25% de ítems en la muestra de reserva) para evaluar el modelo con el conjunto de datos de entrenamiento.
- Calcula y muestra la media de las puntuaciones AUC obtenidas durante la validación cruzada.
- Obtén las predicciones sobre el conjunto de prueba.
- Obtén el informe de clasificación del conjunto de prueba, utilizando la librería sklearn.
- Obtén la curva ROC y el área bajo la misma (AUC).

Practica con Otros Datasets

- Repite la práctica con otros datasets disponibles en el UCI Machine Learning Repository y ya conocidos tanto de Regresión como de Clasificación. Por ejemplo:
 - ▶ **Pima Indians Diabetes Database:**
<https://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes>
 - ▶ **Heart Disease Dataset:**
<https://archive.ics.uci.edu/ml/datasets/Heart+Disease>
 - ▶ **Bank Marketing Dataset:**
<https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>
 - ▶ **Adult (Census Income) Dataset:**
<https://archive.ics.uci.edu/ml/datasets/Adult>
 - ▶ **Auto MPG Dataset:** <https://archive.ics.uci.edu/dataset/9/auto+mpg>
 - ▶ **Wine Quality Dataset:**
<https://archive.ics.uci.edu/dataset/186/wine+quality>
 - ▶ **Energy Efficiency Dataset:**
<https://archive.ics.uci.edu/dataset/242/energy+efficiency>
 - ▶ **Concrete Compressive Strength:**
<https://archive.ics.uci.edu/dataset/165/concrete+compressive+strength>
- Sigue el mismo flujo de trabajo: carga los datos, explora las variables, preprocesa el dataset, ajusta el modelo y evalúa los resultados sin y con validación cruzada. Prueba también a seleccionar prototipos.
- Compara los resultados con aquellos obtenidos con Regresión Lineal, Logística y/o kNN.



UNIVERSITAS
Miguel Hernández

Tema 5. Árboles de Decisión. Práctica.

José L. Sainz-Pardo Auñón

TÉCNICAS ESTADÍSTICAS PARA EL APRENDIZAJE II

Máster Universitario en Estadística Computacional
y Ciencia de Datos para la Toma de Decisiones.

Descarga y Carga de los Datos

- Descarga el dataset **Breast Cancer Wisconsin (Diagnostic)** desde el UCI Machine Learning Repository: [https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)).
- Carga los datos en tu entorno de programación utilizando Python. Asegúrate de que todas las variables se carguen correctamente.
- Revisa las primeras filas del dataset y obtén un resumen estadístico básico para entender la distribución de las características.

Explora los Datos

- Analiza la variable objetivo (Diagnosis) y asegúrate de identificar la proporción de casos benignos y malignos.
- Visualiza las distribuciones de las variables más importantes como Radius Mean, Texture Mean, y otras características clave.
- Convierte la variable Diagnosis a numérica: asigna 0 a los valores "B" (benigno) y 1 a los valores "M" (maligno).
- Dibuja un mapa de correlaciones entre las variables para identificar relaciones importantes.

Preprocesa los Datos

- Elimina la columna ID, ya que no aporta información relevante al modelo.
- Divide el dataset en conjunto de entrenamiento (70%) y de prueba (30%) utilizando la función `train_test_split`.

Ajusta el Modelo de Arboles de Decisión

- Obtén varios árboles de decisión configurando distintos parámetros (prueba manualmente y con un grid de parámetros de búsqueda).
- Visualiza sus esquemas de decisión. Interpretalos. ¿Qué árbol es el que consideras más relevante?
- Obtén los pronósticos de la muestra de prueba con los parámetros del árbol que consideres más relevante.
- Obtén la tabla de confusión y el informe de clasificación del modelo, utilizando la librería sklearn.

Practica con Otros Datasets

- Repite la práctica con otros datasets disponibles en el UCI Machine Learning Repository y ya conocidos tanto de Regresión como de Clasificación. Por ejemplo:
 - ▶ **Pima Indians Diabetes Database:**
<https://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes>
 - ▶ **Heart Disease Dataset:**
<https://archive.ics.uci.edu/ml/datasets/Heart+Disease>
 - ▶ **Bank Marketing Dataset:**
<https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>
 - ▶ **Adult (Census Income) Dataset:**
<https://archive.ics.uci.edu/ml/datasets/Adult>
 - ▶ **Auto MPG Dataset:** <https://archive.ics.uci.edu/dataset/9/auto+mpg>
 - ▶ **Wine Quality Dataset:**
<https://archive.ics.uci.edu/dataset/186/wine+quality>
 - ▶ **Energy Efficiency Dataset:**
<https://archive.ics.uci.edu/dataset/242/energy+efficiency>
 - ▶ **Concrete Compressive Strength:**
<https://archive.ics.uci.edu/dataset/165/concrete+compressive+strength>
- Sigue el mismo flujo de trabajo: carga los datos, explora las variables, preprocesa el dataset, ajusta el modelo y evalúa los resultados sin y con validación cruzada. Prueba también a seleccionar prototipos.
- Compara los resultados con aquellos obtenidos con Regresión Lineal, Logística y/o kNN.



UNIVERSITAS
Miguel Hernández

Tema 6. Random Forest. Prácticas.

José L. Sainz-Pardo Auñón

TÉCNICAS ESTADÍSTICAS PARA EL APRENDIZAJE II

Máster Universitario en Estadística Computacional
y Ciencia de Datos para la Toma de Decisiones.

Descarga y Carga de los Datos

- Descarga el dataset **Breast Cancer Wisconsin (Diagnostic)** desde el UCI Machine Learning Repository: [https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)).
- Carga los datos en tu entorno de programación utilizando Python. Asegúrate de que todas las variables se carguen correctamente.
- Revisa las primeras filas del dataset y obtén un resumen estadístico básico para entender la distribución de las características.

Explora los Datos

- Analiza la variable objetivo (Diagnosis) y asegúrate de identificar la proporción de casos benignos y malignos.
- Visualiza las distribuciones de las variables más importantes como Radius Mean, Texture Mean, y otras características clave.
- Convierte la variable Diagnosis a numérica: asigna 0 a los valores "B" (benigno) y 1 a los valores "M" (maligno).
- Dibuja un mapa de correlaciones entre las variables para identificar relaciones importantes.

Preprocesa los Datos

- Elimina la columna ID, ya que no aporta información relevante al modelo.
- Divide el dataset en conjunto de entrenamiento (70%) y de prueba (30%) utilizando la función `train_test_split`.

Ajusta el Modelo de Árboles de Decisión

- Obtén varios árboles de decisión configurando distintos parámetros (prueba manualmente y con un grid de parámetros de búsqueda).
- Visualiza sus esquemas de decisión. Interprétalos. ¿Qué árbol es el que consideras más relevante?
- Obtén los pronósticos de la muestra de prueba con los parámetros del árbol que consideres más relevante.
- Obtén la tabla de confusión y el informe de clasificación del modelo, utilizando la librería sklearn.

Practica con Otros Datasets

- Repite la práctica con otros datasets disponibles en el UCI Machine Learning Repository y ya conocidos tanto de Regresión como de Clasificación. Por ejemplo:
 - ▶ **Pima Indians Diabetes Database:**
<https://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes>
 - ▶ **Heart Disease Dataset:**
<https://archive.ics.uci.edu/ml/datasets/Heart+Disease>
 - ▶ **Bank Marketing Dataset:**
<https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>
 - ▶ **Adult (Census Income) Dataset:**
<https://archive.ics.uci.edu/ml/datasets/Adult>
 - ▶ **Auto MPG Dataset:** <https://archive.ics.uci.edu/dataset/9/auto+mpg>
 - ▶ **Wine Quality Dataset:**
<https://archive.ics.uci.edu/dataset/186/wine+quality>
 - ▶ **Energy Efficiency Dataset:**
<https://archive.ics.uci.edu/dataset/242/energy+efficiency>
 - ▶ **Concrete Compressive Strength:**
<https://archive.ics.uci.edu/dataset/165/concrete+compressive+strength>
- Sigue el mismo flujo de trabajo: carga los datos, explora las variables, preprocesa el dataset, ajusta el modelo y evalúa los resultados sin y con validación cruzada. Prueba también a seleccionar prototipos.
- Compara los resultados con aquellos obtenidos con Regresión Lineal, Logística y/o kNN.