APRENDIZAJE ESTADÍSTICO

Boletín 1: Evaluación y Selección de Modelos (KNN)

**ANDRÉS CAMPOS CUIÑA**

**FECHA DE ENTREGA: 19/11/2021**

ÍNDICE

[1 Ejercicio 1 1](#_Toc88044118)

[2 Ejercicio 2 2](#_Toc88044119)

[3 Ejercicio 3 12](#_Toc88044120)

# Ejercicio 1

**Suponiendo que se quiere hacer la predicción de la variable de salida para X1=0, X2=0, X3=0 mediante KNN.**

**a. Computar la distancia entre cada observación y el punto de test.**

Distancia entre el punto [0 3 2] y el punto de test [0 0 0] = 3.605551275463989

Distancia entre el punto [3 0 3] y el punto de test [0 0 0] = 4.242640687119285

Distancia entre el punto [ 0 3 -1] y el punto de test [0 0 0] = 3.1622776601683795

Distancia entre el punto [3 0 0] y el punto de test [0 0 0] = 3.0

Distancia entre el punto [1 2 1] y el punto de test [0 0 0] = 2.449489742783178

Distancia entre el punto [2 1 0] y el punto de test [0 0 0] = 2.23606797749979

**b. ¿Cuál es la predicción para K=1? ¿Por qué?**

La predicción para [0 0 0] es 0 y el 1 vecino más cercanos es:

El vecino [2 1 0] a una distancia de 2.23606797749979 con clase 0

**c. ¿Cuál es la predicción para K=3? ¿Por qué?**

La predicción para [0 0 0] es 1 y los 3 vecino más cercanos son:

El vecino [2 1 0] a una distancia de 2.23606797749979 con clase 0

El vecino [1 2 1] a una distancia de 2.449489742783178 con clase 1

El vecino [3 0 0] a una distancia de 3.0 con clase 1

# Ejercicio 2

**Dado el problema de clasificación Blood Transfusion Service Center:**

**a. Analiza las características del conjunto de datos: número y tipo de variables de entrada y salida, número de instancias, número de clases y distribución de las mismas, correlación entre las variables, valores perdidos, etc.**

Tabla

Descripción generada automáticamente con confianza media

**Tipo de dato de cada columna del Dataframe :**

Recency: int64

Frequency: int64

Monetary: int64

Time: int64

IsMarchDonor: int64

**Número de clases y distribución de las mismas:**

Gráfico

Descripción generada automáticamente

Gráfico

Descripción generada automáticamente

Gráfico

Descripción generada automáticamente

Gráfico, Gráfico de barras

Descripción generada automáticamente

Captura de pantalla de computadora

Descripción generada automáticamente con confianza media

**La correlación entre las variables:**

Imagen que contiene Gráfico

Descripción generada automáticamente

Imagen que contiene Gráfico

Descripción generada automáticamente

Imagen que contiene Gráfico

Descripción generada automáticamente

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

Calendario

Descripción generada automáticamente con confianza baja

Calendario

Descripción generada automáticamente con confianza baja

**b. Una de las clases que implementa el algoritmo KNN en *scikit‐learn* es *sklearn.neighbors.KNeighborsClassifier*. Revisa los parámetros y métodos que tiene.**

Revisado.

**c. Divide los datos en entrenamiento (80%) y test (20%).**

Hecho.

**d. Realiza la experimentación con KNN (*KNeighborsClassifier*) usando como hiper‐parámetro el número de vecinos.**

**Muestra la gráfica del error de entrenamiento con validación cruzada (5‐CV) frente al valor del hiper‐parámetro. ¿Cuál es el menor error de validación cruzada, su desviación estándar y el valor del hiper‐parámetro para el que se consigue? ¿Cuál es el valor del hiperparámetro si se aplicase la regla de una desviación estándar?**

**Gráfico, Gráfico de líneas

Descripción generada automáticamenteMenor error de validación cruzada, su desviación estándar y el valor del hiper‐parámetro para el que se consigue:**

**Imagen que contiene Tabla

Descripción generada automáticamente**

El error real es: 0.2040896358543417

**El valor del hiper-parámetro si se aplicase la regla de una desviación estándar:**

param\_n\_neighbors: 55

mean\_test\_score: 0.772577

std\_test\_score: 0.009657

rank\_test\_score: 41

Error real: 0.22742296918767513

**La gráfica de la selección de este valor es la siguiente:**

Gráfico, Gráfico de líneas

Descripción generada automáticamente

**Muestra la gráfica del error de test frente al valor del hiper‐parámetro, y valora si la gráfica del error de entrenamiento con validación cruzada ha hecho una buena estimación del error de test. ¿Cuál es el menor error de test y el valor del hiper‐parámetro para el que se consigue? ¿Cuál es el error de test para el valor del hiper‐parámetro seleccionado por la validación cruzada? ¿Cuál es el error de test para el valor del hiper‐parámetro seleccionado por la validación cruzada mediante la regla de una desviación estándar?**

**Gráfico

Descripción generada automáticamente**

**El menor error de test y el valor del hiper‐parámetro para el que se consigue:**

**Texto

Descripción generada automáticamente con confianza baja**

Error real: 0.19333333333333336

**El error de test para el valor del hiper‐parámetro seleccionado por la validación cruzada:**

0.21333333333333337

**El error de test para el valor del hiper‐parámetro seleccionado por la validación cruzada mediante la regla de una desviación estándar:**

0.22666666666666668

# Ejercicio 3

**Dado el problema de regresión Energy Efficiency:**

**a. Analiza las características del conjunto de datos: número y tipo de variables de entrada y salida, número de instancias, número de clases y distribución de las mismas, correlación entre las variables, valores perdidos, etc.**

Tabla

Descripción generada automáticamente

**Tipo de dato de cada columna del Dataframe :**

X1 float64

X2 float64

X3 float64

X4 float64

X5 float64

X6 int64

X7 float64

X8 int64

Y float64

**Número de clases y distribución de las mismas:**

Gráfico, Histograma, Gráfico en cascada

Descripción generada automáticamente

Gráfico, Histograma

Descripción generada automáticamente

Gráfico, Gráfico de barras, Histograma

Descripción generada automáticamente

**La correlación entre las variables:**

Imagen que contiene luz, computadora, tabla, computer

Descripción generada automáticamente

Calendario

Descripción generada automáticamente

**b. Una de las clases que implementa el algoritmo KNN en *scikit‐learn* es *sklearn.neighbors.KNeighborsRegressor*. Revisa los parámetros y métodos que tiene.**

Revisado.

**c. Divide los datos en entrenamiento (80%) y test (20%).**

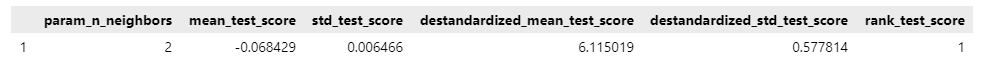
Hecho.

**d. Realiza la experimentación con KNN (*KNeighborsRegressor*) usando como hiper‐parámetro el número de vecinos.**

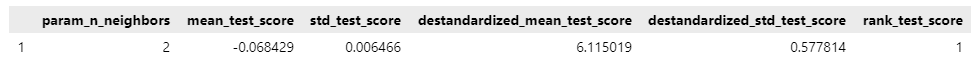
**Muestra la gráfica del error de entrenamiento con validación cruzada (5‐CV) frente al valor del hiper‐parámetro. ¿Cuál es el menor error de validación cruzada, su desviación estándar y el valor del hiper‐parámetro para el que se consigue? ¿Cuál es el valor del hiperparámetro si se aplicase la regla de una desviación estándar?**

**Gráfico, Gráfico de líneas

Descripción generada automáticamenteMenor error de validación cruzada, su desviación estándar y el valor del hiper‐parámetro para el que se consigue:**

****

**El valor del hiper-parámetro si se aplicase la regla de una desviación estándar:**

****

**La gráfica de la selección de este valor es la siguiente:**

Gráfico

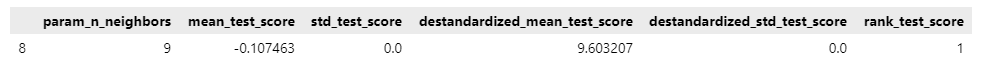
Descripción generada automáticamente

**Muestra la gráfica del error de test frente al valor del hiper‐parámetro, y valora si la gráfica del error de entrenamiento con validación cruzada ha hecho una buena estimación del error de test. ¿Cuál es el menor error de test y el valor del hiper‐parámetro para el que se consigue? ¿Cuál es el error de test para el valor del hiper‐parámetro seleccionado por la validación cruzada? ¿Cuál es el error de test para el valor del hiper‐parámetro seleccionado por la validación cruzada mediante la regla de una desviación estándar?**

**Gráfico, Gráfico de líneas, Histograma

Descripción generada automáticamente**

**El menor error de test y el valor del hiper‐parámetro para el que se consigue:**



**El error de test para el valor del hiper‐parámetro seleccionado por la validación cruzada:**

12.588656331168831

**El error de test para el valor del hiper‐parámetro seleccionado por la validación cruzada mediante la regla de una desviación estándar:**

12.588656331168831