

Ejercicios Aprendizaje Automático.

Instrucciones.

La solución de los ejercicios se ha de entregar en un archivo en formato IPython Notebook. Este archivo ha de contener tanto el código necesario para la resolución de los problemas como los comentarios en los que explican los resultados. Los archivos necesarios para la solución de los problemas se han de cargar desde la misma carpeta en la que encuentre el Notebook utilizando siempre rutas relativas.

Cada ejercicio tiene una puntuación diferente que se muestra en su enunciado.

1. Reglas de asociación. (10 puntos)

En el “groceries.csv” contiene las transacciones recogidas durante un mes en una tienda de comestibles. En cada una de las filas de este archivo se encuentran los artículos comprados de forma conjunta por los diferentes clientes. Obtener la lista de artículos que tiene un soporte mínimo de 0,15. Obtener también las reglas de asociación que se pueden deducir de este conjunto de datos con un soporte mínimo de 0,05 y confianza de 0,25.

La importación del archivo se puede realizar utilizando el siguiente código:

```
import csv

groceries = []
groceries_file = csv.reader(open("groceries.csv", "rb"))

for row in groceries_file:
    groceries.append(row)
```

2. Implementación de una regresión “stepwise” con eliminación hacia atrás. (40 puntos)

Utilizando como referencia el código del algoritmo “stepwise” con selección hacia adelante (Forward Stepwise Regression) que se encuentra en el archivo “CIF005_02_06_Stepwise.ipynb” realizar una implementación del algoritmo con eliminación hacia atrás (Backward Stepwise Regression). En este caso la selección de las variables se realiza empezando con un modelo que utiliza todas las variables disponibles para ir eliminando en cada paso la que produce el modelo menos significativo.

3. Estimación de la supervivencia de los pasajeros del Titanic. (50 puntos)

Utilizando los datos del archivo “titanic.csv”, que contienen información del pasaje del Titanic, para la creación de un modelo que pueda predecir si un pasajero sobrevive al hundimiento del barco en el momento de embarque. Para esto se han de seleccionar las variables con mayor capacidad de predicción.

Para la creación del modelo no se pueden utilizar las variables “boat” ni “body” ya que estas contienen información posterior al momento del embarque.

En el conjunto de datos existes valores NaN en las columnas “age” y “fare”, para eliminar las filas con estos valores se puede utilizar el siguiente comando:

```
mask = titanic[['age', 'fare']].applymap(lambda x: math.isnan(x))
titanic = titanic[-mask.any(axis = 1)]
```