

Clase10 IMA539

Eduardo Castro Thompson

April 21, 2023

Clase10 IMA539

Sobre la base de datos: *creditcard.csv*. - Contiene información sobre **284,807 transacciones de las cuales 492 son consideradas fraudulentas**. - Las características contenidas son las siguientes: - 'V1' a 'V28' son las características obtenidas luego de una transformación utilizando PCA. - 'Time' contiene la cantidad de tiempo pasada entre la transacción correspondiente y la primera transacción del set. - 'Amount' contiene el valor de la transacción efectuada. - 'Class' contiene las etiquetas: 1 para fraudes y 0 en otro caso.

Ejercicios propuestos:

1. Lea el archivo local `creditcard.csv`
2. Genere una base de datos “de paso” que contenga 10.000 transacciones legítimas y todas las fraudulentas, luego, utilice TSNE para visualizar los clusters encontrados.
3. Instancie dos modelos: En primer lugar, una pipeline que realice una reducción de dimensión a 15 características utilizando PCA seguido de un Perceptrón y por otro lado un modelo Random Forest.
4. Estime el desempeño de los modelos utilizando Cross Validation. Reporte los valores de Exactitud, Precisión, Recall, F1-Score, AUC, MCC y Kappa. ¿Son confiables estos resultados? ¿Tienen sentido?
5. Balancee los datos y repita los pasos 3 y 4 ¿Ha cambiado el valor de las métricas? Si es así ¿Por qué?
6. Muestre la curva de aprendizaje de los modelos en los datos balanceados con respecto al F1-Score.

0.0.1 Librerías

```
[1]: # Metrics and analysis
from sklearn.metrics import roc_auc_score, matthews_corrcoef, \
    cohen_kappa_score, make_scorer
from sklearn.model_selection import learning_curve

# Data modelation
from sklearn.model_selection import cross_validate
```

```
# Classifiers
from sklearn.ensemble import RandomForestClassifier
from sklearn.linear_model import Perceptron

# Dimension reduction
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.manifold import TSNE

# Pipeline
from sklearn.pipeline import make_pipeline

import matplotlib.patheffects as PathEffects
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
```

0.0.2 Desarrollo

[]: