

Curso

Machine Learning con Python

Unidades 5-8

Trabajo Práctico Final



Objetivos

Aplicar los conocimientos sobre aprendizaje supervisado y no supervisado



Consignas

Utilizando los datos de vinos del conjunto de datos “Wine recognition dataset” de Scikit-Learn, que puede obtener con la instrucción

```
from sklearn.datasets import load_wine
```

- 1) Construya 2 datasets, uno con los datos originales y otro con los datos escalados y/o normalizados. Divida ambos datasets en conjuntos de entrenamiento y de validación.
- 2) Elija un valor de K y aplique el algoritmo de KNN en ambos datasets. Evalúe la precisión en ambos casos sobre el conjunto de validación correspondiente. Analice brevemente los resultados obtenidos
- 3) Elija un valor de max_depth y aplique el algoritmo de Árboles de Decisión en ambos datasets. Evalúe la precisión en ambos casos. sobre el conjunto de validación correspondiente. Analice brevemente los resultados obtenidos

- 4) Aplique el algoritmo PCA a los datos escalados y/o normalizados para reducir su dimensionalidad a 2, y grafique el conjunto resultante.
- 5) Sin utilizar las etiquetas del dataset, elija un valor de K y aplique el algoritmo de KMeans sobre los datos transformados por PCA.
- 6) Asigne las etiquetas correspondientes a los datos de dimensionalidad reducida obtenidos en 4), y divida dichos datos en conjuntos de entrenamiento y validación. Elija un valor de K y aplique el algoritmo de KNN. Evalúe la precisión sobre el conjunto de validación. Analice brevemente los resultados obtenidos

NOTA: Puede obtener más información sobre el dataset en

https://scikit-learn.org/stable/datasets/toy_dataset.html#wine-dataset

Formato de presentación:

- El trabajo puede realizarse individualmente, o en grupos de 2 participantes
- Los participantes deben entregar un archivo comprimido que contenga el jupyter notebook desarrollado. El nombre del archivo comprimido debe tener el formato TP_Apellidos. Por ejemplo, TP_Canelon o TP_Barreto_Canelon. El nombre del Jupyter notebook debe tener el formato Apellidos_TPF.ipynb. Por ejemplo, Canelon_TPF.ipynb o Barreto_Canelon_TPF.ipynb
- El archivo comprimido debe generarse con la aplicación 7-zip, que puede descargarse gratuitamente desde <https://www.7-zip.org/>.

Fecha límite de entrega:

Nominal: 08/05/2024 - 23:59 hrs

Recuperatorio: 15/05/2024 - 23:59 hrs

Las fechas de entrega son inapelables, ya que están configuradas automáticamente en el Campus. La plataforma no permitirá que los participantes entreguen fuera de la fecha/hora indicada. Si no pueden hacerlo en la primera, podrán hacerlo en el recuperatorio.

Criterios de evaluación

- La calificación total del trabajo estará en función del número de consignas realizadas correctamente. Si alguna consigna no funciona de manera correcta o genera un error en el Jupyter notebook, se restarán puntos del total correspondiente a esa consigna



Bibliografía utilizada y sugerida

scikit-learn. Documentación oficial.

Disponible desde: URL: <https://scikit-learn.org/stable/>

NumPy. Documentación oficial.

Disponible desde: URL: <https://numpy.org/>

Pandas. Documentación oficial.

Disponible desde: URL: <https://pandas.pydata.org/>

Matplotlib. Documentación oficial.

Disponible desde: URL: <https://matplotlib.org/>