

Curso de Machine Learning

Unidad Nro 8

Trabajo Práctico Final



Objetivos

Poner en práctica los conceptos de aprendizaje automático supervisado y no supervisado.
Familiarizarse con la adquisición, inspección y visualización de datos.
Implementar, entrenar y validar los modelos vistos en clase.



Consigna

Aprendizaje supervisado y no supervisado.

Utilizando los datos de vinos obtenidos del siguiente dataset:

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

- 1) Construya 2 datasets, uno con los datos originales y otro con los datos correctamente 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.
- 3) Elija un valor de max_depth y aplique el algoritmo de Árboles de Decisión en ambos datasets. Evalúe la precisión en ambos casos.
- 4) Tome los datos escalados y/o normalizados, reduzca la dimensión del problema a 2 utilizando PCA 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. Repita el procedimiento utilizando KMeans por mini-batches.

IMPORTANTE: Analice los resultados de los puntos anteriores brevemente.



Formato de presentación:

- Un solo archivo comprimido con deberá contener la(s) jupyter notebooks **.ipynb** que ejecuten el código de todos los modelos.



Fecha límite de entrega:

Nominal: 25/01/2023 - 00:00 hrs

Recuperatorio: 01/02/2023 - 00:00 hrs



Criterios de evaluación

Para la calificación de la tarea tendremos en cuenta el procedimiento de preparación, inspección y visualización de datos, así como la correcta implementación, entrenamiento y validación de los modelos. Se priorizará el cumplimiento de todos los incisos de las consignas.

La eficiencia del código no se considerará un factor determinante de la calificación final, pero se darán puntos extras por la utilización de herramientas de numpy, matplotlib, sklearn, etc...



Bibliografía utilizada y sugerida

Documentación

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/>