

PLE01

Aprendizaje Automático

Borja González Seoane

S02, 11 de septiembre de 2024

1. Descripción de la práctica

La primera de las prácticas de laboratorio evaluada, PLE01, de la asignatura de Aprendizaje Automático consistirá en la realización y presentación de un análisis exploratorio de datos (EDA). Cada estudiante deberá escoger un conjunto de datos del repositorio UCI Machine Learning Repository¹ para proceder con el análisis.

Para emprender la tarea, el estudiante deberá emplear las herramientas utilizadas ya en la prácticas de laboratorio de las últimas sesiones: PL01 y PL02. En concreto, se deberá emplear el lenguaje de programación Python y las librerías que ya introducidas en susodichas prácticas, como Pandas, Matplotlib, Seaborn, Plotly, NumPy, Scikit-learn, etc.

Se espera que el estudiante cree un Jupyter Notebook en el que se realice el análisis exploratorio de datos. Luego, opcionalmente —aunque en principio recomendable—, el estudiante podrá extraer los resultados, visualizaciones o conclusiones más relevantes del EDA para generar unas transparencias en las que apoyarse durante la presentación oral de la práctica. Sea cual sea el soporte empleado, el estudiante deberá presentar los resultados de su análisis exploratorio de datos en una breve exposición de unos 10 minutos, que deberá abarcar:

- Una descripción general del conjunto de datos: contexto, para qué se podría emplear (etiqueta objetivo)...
- Descripción detallada de las variables del conjunto de datos: tipos, rangos, valores nulos, valores atípicos, distribuciones de las variables, etc.
- Estadísticas descriptivas de las variables del conjunto de datos. Visualizaciones con las que poder explicar la naturaleza de las variables del conjunto de datos.
- Recortes del código Python empleado para realizar el análisis. Partes más relevantes. Se pueden llevar recortes a las transparencias o ir cambiando de las transparencias al *notebook* durante la exposición.
- Conclusiones finales del análisis exploratorio de datos. Resumen de los hallazgos más relevantes del conjunto de datos.

Dada la naturaleza de la práctica, se recomienda que el estudiante escoja un conjunto de datos que le resulte interesante y que le motive a realizar un análisis exhaustivo. Se recomienda también que el estudiante se centre en la calidad del análisis exploratorio de datos, más que en la cantidad de visualizaciones o estadísticas descriptivas que pueda generar.

¹<https://archive.ics.uci.edu/ml/index.php>

Además de lo anterior, cada estudiante deberá elegir un conjunto de datos diferente, para fomentar la diversidad de análisis y la variedad de conjuntos de datos empleados. Para ello, se habilitará en el Campus Virtual un hilo de discusión en el que los estudiantes podrán ir indicando qué conjunto de datos han escogido, para que el resto de compañeros pueda elegir uno distinto.

2. Entrega y evaluación de la práctica

La entrega de la práctica se realizará a través del Campus Virtual, se habilitará una tarea específica para ello. Los entregables de la tarea serían el Jupyter Notebook del EDA y el PDF de las transparencias de la presentación, en caso de que se haya decidido prepararlas.

La práctica será expuesta oralmente en la sesión S03 del curso. La tarea para subir los entregables se cerrará una hora después de la conclusión de la sesión.

La evaluación de la práctica se realizará basándose en los siguientes criterios:

- Calidad del análisis exploratorio de datos: profundidad, rigor, relevancia de los hallazgos, etc.
- Calidad de la presentación: claridad, estructura, calidad de las visualizaciones, etc.
- Uso de las herramientas de programación: calidad del código, uso de las librerías adecuadas, etc.
- Ajuste al tiempo de exposición. Máximo de 10 minutos.

Se recuerda que la calificación de esta práctica supondrá un 10% de la calificación global de la asignatura, tal y como figura en el Contrato de Enseñanza-Aprendizaje.