Unidad 3 - Trabajo Práctico
Preprocesamiento y análisis preliminar de datos

Enunciado

El objetivo de este trabajo práctico es aprender a reconocer problemas de regresión y diseñar/codificar algoritmos para su resolución.

Para ello, les pedimos que:

- Tomen alguna de las bases de datos provistas por el curso (disponibles en https://github.com/ignaciorlando/duia-ml-datasets) que se asocie a un problema de regresión (tienen que identificarlo uds mismos!).
- Describan el problema de regresión que reconocen, las características involucradas para cada muestra y la variable respuesta que quieren predecir.
- Realicen una partición en datos de entrenamiento, validación y test, aplicando las técnicas de estandarización que aprendimos en las unidades anteriores. Tengan en cuenta que pueden usar las rutinas de sklearn.model_selection para facilitar el particionado, no hace falta que implementen todo el proceso con Python puro que se hizo en el Colab de ejemplo de la Unidad 1.
- Seleccionen dos de los modelos de regresión lineal que vimos en la teoría (Regresión lineal estándar, Ridge Regression, LASSO o Random Forest Regression) y reproduzcan el pipeline completo de entrenamiento/validación/test. Para ello:
    - Realizar model selection para optimizar los hiperparámetros de los dos algoritmos.
    - Evaluar los modelos óptimos sobre los datos de test, utilizando algunas de las métricas de evaluación vistas en la materia, y comparar los resultados obtenidos.

Entrega

Un Colab que documente el análisis de los datos y todo el proceso de resolución del problema.

Condiciones de entrega y aprobación

Para la evaluación de la materia pueden entregar un solo práctico de los que les daremos para la Unidad 3, 4 y 5. Sí o sí deben entregar alguno de los 3, pero no es condición necesaria para aprobar entregarlos a todos.

- El trabajo práctico puede hacerse en grupos de hasta 2 personas.
- No puede utilizarse el mismo data set que se usó durante la clase.
- Asegurarse de que el código incluya soporte para descargar los datos, o incluir los datos en el repositorio.

    - Si en el trabajo práctico de la Unidad 1 utilizaron un data set que sirve para resolver un problema de regresión, pueden utilizarlo. Tengan en cuenta que deberán copiarse las rutinas para preprocesarlo, o bien implementar las rutinas para descargarlo automáticamente desde un link que uds generen.
- La entrega se realiza a través de Classroom. Incluyan por favor el nombre y apellido completo de los miembros del grupo en el Colab que hagan.