

Taller de Aprendizaje Automático

Instituto de Ingeniería Eléctrica - FING - UdelAR

Entrega 1 – 2024

- Esta entrega cubre las actividades realizadas en los Talleres 3, 4 y 5.
- Se debe escribir un informe de no más de **cuatro carillas**.
- El trabajo es **individual**. Se valora especialmente la originalidad de los abordajes y soluciones descriptas, en especial entre miembros de un grupo.

Talleres 3 y 5 – *Demanda de bicicletas*

1. Considere la determinación del valor óptimo del hiperparámetro utilizado para regularizar el árbol. Indique el rango de valores y la estrategia utilizados en la búsqueda. Ilustre gráficamente.
2. Desde el punto de vista teórico ¿por qué la combinación de árboles puede mejorar el desempeño de uno individual? ¿qué indican sus resultados al respecto?
3. Compare los resultados obtenidos en *Kaggle* con el mejor estimador del taller 3 y el mejor estimador del taller 5. Describa cada uno de los *pipelines*. En caso de haber realizado alguna parte opcional que no haya sido la que produjo los mejores resultados, agréguela en la comparación y explique por qué consideró que podía ser una opción interesante a explorar.
4. Utilizando la técnica de *shapley values* explique la predicción de alquileres realizada por su mejor modelo para el día 24 de abril de 2011 a las 5 de la tarde.
5. Junto al informe suba el mejor modelo obtenido con un código que permita entrenar y correr el modelo ya entrenado (notebook o script de python).

Taller 4 – *Detección de anomalías*

1. ¿Le parece razonable abordar el problema como uno de detección de anomalías? ¿Por qué?
2. Comente sobre el *pipeline* de preprocesamiento utilizado, explicando por qué incluyo cada uno de los componentes.
3. Compare la detección de anomalías utilizando al menos dos estrategias a través de sus curvas *precision-recall*.

4. Elija una de las estrategias, determine su punto de operación. Comente las razones que motivaron su decisión.
5. ¿Qué resultados obtuvo para los datos de test? ¿Están dentro de lo esperado? Se sugiere utilizar la tasa de aciertos por categoría de anomalía. ¿Cuáles son detectadas correctamente?