

Taller de Aprendizaje Automático

Instituto de Ingeniería Eléctrica - FING - UdelaR

Proyecto 1 – 2024

Bosón de Higgs

Descripción

El proyecto consiste en diseñar e implementar un clasificador para el problema de detección del bosón de Higgs. Se trabaja en grupos de tres estudiantes.

Objetivos de aprendizaje

Para la resolución del proyecto debería ser capaz de realizar lo siguiente:

1. Explorar y visualizar los datos e identificar características relevantes.
2. Implementar pipelines en *Scikit-learn* y funciones para conseguir los datos.
3. Limpiar y pre-procesar los datos. Manejar datos faltantes.
4. Implementar, entrenar y refinar modelos de clasificación y regresión.
5. Aplicar técnicas de combinación de clasificadores débiles.
6. Manejar adecuadamente conjuntos de entrenamiento, validación y prueba.
7. Manejar el desbalance de clases y el peso relativo de los datos.
8. Identificar el sobreajuste y sub-ajuste de un modelo a los datos.
9. Aplicar técnicas de regularización y de selección de hiperparámetros.
10. Comparar soluciones usando diferentes medidas de desempeño.
11. Elegir la solución final, incluyendo la selección del punto de trabajo.
12. Aplicar alguna técnica que permita explicar las predicciones realizadas por el clasificador, por ejemplo *Shapley values*.
13. Diseñar y registrar experimentos usando una herramienta como *Comet*.
14. Usar la plataforma *Kaggle* para obtener los datos y subir la solución final.

Informe y entrega

El documento no debe exceder las **8 páginas**. El informe debe presentar muy brevemente el problema abordado y describir en detalle la solución desarrollada. Se deben justificar las decisiones tomadas en el diseño de la solución, en particular respecto al manejo de los datos (preprocesamiento, desbalance), la comparación entre técnicas de clasificación (al menos dos), la regularización y la selección de hiperparámetros.

Deben incluir en la entrega (junto al informe) el modelo final con un código que permita ejecutarlo sobre los datos de test. Además, deben entregar un registro de los principales experimentos realizados (link a *Comet*) junto a código usado para generarlos.

Recuerden hacer referencia en el texto a las figuras que incluyan, y que cada figura cuente con una nota de descripción. Tengan la precaución de que las etiquetas y números de las figuras sean legibles. Puede ser conveniente resumir los resultados de desempeño usando tablas. Incluyan una sección de referencias bibliográficas, y citen cuando crean necesario.

Defensa

Cada grupo de estudiantes tendrá **10 minutos** para hacer una presentación oral de su trabajo (seremos estrictos en cuanto al tiempo) seguido de una sesión de preguntas de unos **20 minutos**. En la defensa el grupo deberá presentar su solución al problema planteado, así como el proceso que condujo a ella, los principales experimentos realizados haciendo énfasis en el análisis de resultados. Se valorará el análisis de la solución vinculado con los conceptos y técnicas vistas en el curso. No es necesario describir el problema en la presentación.

Fechas

- Entrega de informe: martes 7 de mayo, 13:00 hrs.
- Defensas: jueves 9 de mayo de 8 a 12 y de 13 a 15 hrs.