

Instrucciones Generales

- La presente primera Tarea será de programación/entrenamiento a partir de un set de datos
- Esta Tarea debe realizarse en forma individual
- Se puede implementar en la plataforma que desee, se sugiere utilizar Python/Jupyter para una corrección más rápida.

Detalles de la Entrega

- Plazo de Entrega: lunes, 11 de Octubre, 23:59hrs
- Subir el documento (p.ej.: t1_apellido_p.ipynb) a la plataforma Canvas en Tareas. "p" corresponde al número de pregunta: $p \in \{1,2\}$.
- El código debe ejecutar correctamente

Contexto

Como fuera expuesto en clases, se trabajará sobre un conjunto de datos etiquetados (etiquetas binarias $\{0,1\}$) y de entradas multivariantes.

La idea es generar diferentes modelos basados en aprendizaje, que implementen una predicción.

Una buena fuente de datos es posible encontrarla en el sitio de desafíos/competencia Kaggle (<https://www.kaggle.com/datasets>)

Preguntas

- 1) Cargar los datos del set (conjunto) seleccionado. Se sugiere usar pandas (usualmente los datos vienen en archivos csv)
 - a. Hacer una breve descripción de estos datos (estadísticas, histogramas, etc.). Se puede utilizar el mismo pandas para este fin.
 - b. Separar los datos en conjunto de entrenamiento y conjunto de pruebas (80%, 20%).
 - c. Describir estos 2 conjuntos de datos. Idear alguna forma de verificar que los datos están balanceados en términos estadísticos (las 2 poblaciones tienen características estadísticas similares)
- 2) Entrenar un sistema de regresión logística. Se sugiere utilizar LogisticRegression de scikit learn.
 - a. Probar diferentes combinaciones de solvers, cantidad de iteraciones
 - b. Mostrar los diferentes resultados al cambiar esta parametrización
 - c. Comentar los resultados al comparar las diferentes soluciones
- 3) Entrenar un sistema de Red Neuronal Densa Superficial (p.ej., 1 capa oculta) para el predictor
 - a. Probar el sistema con diferentes configuraciones de capas y cantidad de neuronas
 - b. Cambiar la función a minimizar (función objetivo – losses)
 - c. Comentar los resultados al comparar las diferentes soluciones