

# Lista de Ejercicios 3

Machine Learning para CCSS

21 de abril, 2024

## Instrucciones

1. **Temas abordados:** Esta lista de ejercicios se enfoca en los temas: Tree Based Models & Neural Networks.
2. **Formación de grupos:** Se permite la formación de grupos de hasta 5 integrantes.
3. **Puntuación de ejercicios:** La lista contiene 5 ejercicios. Cada ejercicio vale 4 puntos.
4. **Formato de entrega:** La resolución de los ejercicios debe presentarse en un archivo jupyter-notebook con las celdas ejecutadas.
5. **Fecha límite de entrega:** La fecha límite para la entrega es el lunes 29 de abril a las 11:59 pm. Un representante del equipo debe subir su solucionario a la actividad correspondiente en la plataforma Paideia. Los nombres y códigos de todos los participantes deben ser incluidos en el solucionario.

## Pregunta 1

Brinde una explicación detallada del algoritmo que se utiliza para implementar un Árbol de Regresión.

## Pregunta 2

Considere la dataset Carseats para predecir *Sales* como variable cuantitativa (puede encontrar el conjunto de datos aquí). Para ello:

1. Divida los datos en conjuntos de entrenamiento y de prueba.
2. Ajuste un árbol de regresión al conjunto de entrenamiento. Interprete los resultados. ¿Qué valor de MSE se obtiene para el conjunto de prueba?
3. Utilice el método Bagging para analizar estos datos. ¿Qué valor de MSE obtiene para el conjunto de prueba?
4. Utilice el método Random Forest para analizar los datos. Repita el procedimiento descrito en el punto 3. Describa el efecto de  $m$ , el número de variables consideradas en cada división, sobre la tasa de error obtenida. Utilice la función `features_importance()` para determinar qué variables son las más importantes.
5. Utilice el método Gradient Boosting. Repita el procedimiento descrito en el punto 4.

### Pregunta 3.

Aplice los métodos Boosting, Bagging y Random Forest a una dataset de su elección. Asegúrese de dividir los datos en los conjuntos de entrenamiento y de prueba, ajustar los modelos sobre el conjunto de entrenamiento y evaluar su desempeño sobre el conjunto de prueba. ¿Cuán precisos son los resultados si los compara con métodos regresión lineal/logística y/o métodos de Regularización? ¿Cuál de todos los modelos implementados muestra un mejor desempeño?

### Pregunta 4

Implemente una red neuronal con la database Default (puede encontrar el set de datos aquí). Use una capa oculta con 10 unidades y regularización por dropout. Compare el desempeño de clasificación con respecto al de la regresión lineal.

### Pregunta 5

Considere la implementación del modelo de Redes Neuronales Convolucionales (CNN) utilizando la dataset *CIFAR 100* presentada en el notebook de clases. En esta implementación inicial, se alcanzó un Accuracy aproximado de 0.44. Su tarea es desarrollar una nueva implementación que supere este valor de Accuracy. Para lograrlo, se le anima a experimentar con los siguientes aspectos del modelo:

- **Arquitectura del Modelo:** Puede considerar la adición de más capas convolucionales y de pooling, incrementar el número de filtros por capa, o modificar el tamaño del kernel.
- **Regularización:** Experimente con diferentes tasas de dropout para prevenir el sobreajuste.
- **Entrenamiento:** Pruebe con distintos tamaños de batch que puedan influir en la convergencia del modelo durante el entrenamiento.
- **Optimización:** Evalúe el uso de diferentes algoritmos de optimización para verificar si mejoran la precisión del modelo.

La meta es ajustar estos parámetros para superar el Accuracy de 0.44 previamente alcanzado. Documente todas las modificaciones realizadas y los resultados obtenidos.