Práctica Módulo 8  
Aprendizaje Estadístico

short line

Ejercicio 0:

Importe los datos de alquileres de bicicletas disponibles en el archivo **hour.csv** en la carpeta **Data**. realice un breve EDA de los datos.

Ejercicio 1:  
Conserve únicamente variables que tenga sentido conservar bajo un esquema de aprendizaje supervisado donde lo que queremos predecir es la cantidad total de alquileres que se van a presentar ese día. Adicionalmente, aplique ingeniería de atributos, cambiando tipos o generando nuevas variables según considere. Llame al dataset resultante **bike\_clean\_df**.

Ejercicio 2:   
Realice un train-test split de los datos, tenga en cuenta que los datos están asociados a fechas, cómo piensa que esto podría afectar la estrategia de train-test a aplicar?

Ejercicio 3:   
Aplique estrategias de feature selection, empleando las técnicas best subset, forward selection, backward selection y sequential replacement. Y compruebe si con todas converge a la misma selección para cada subconjunto de features y la mejor performance a través de r^2 adj y BIC.

Ejercicio 4:   
Al momento está conforme con el poder predictivo en entrenamiento? Compare con una regresión múltiple sobre todos los datos sin ingeniería de atributos.

Ejercicio 5:   
Evalúe el mejor modelo en test usando el RMSE. y grafique en un gráfico de dispersión los resultados reales vs los estimados.

Ejercicio 6:   
Investigue en la documentación de la implementación de algún algoritmo no desarrollado en el curso (PC Regression, árboles de decisión, Random Forests, XGBoost, etc) e intente usarlo para armar un modelo con el conjunto de entrenamiento y ponerlo a prueba con el set de testeo. Comparar RMSE y graficar predicciones. (difícil)