

Taller de modelado matemático I

Tarea 2

Esta es la segunda entrega de 3 tareas que van encaminadas en generar un modelo predictivo para predecir el precio final de cada casa de un conjunto de datos con variables explicativas que describen aspectos de las casas residenciales en Ames, Iowa. Esta serie de tareas **PUEDE** ser en equipo, de 3 máximo personas. El modelo final se tiene que ajustar o entrenar en los datos Casas.csv (con aproximadamente 1400 registros) y probar su poder predictivo en los datos Casas_Kaggle.csv (con aproximadamente 1400 registros). La descripción de las variables está en el archivo data_description.txt. El **equipo o persona** que al final genere el modelo con el mejor poder predictivo tendrá 10% extra en las tareas. Para evaluar los resultados de la predicción, tienen que entrar a la página de kaggle (previamente registrarse e iniciar una sesión). En la tarea 3 se explicará cómo se somete la predicción.

La segunda tarea consiste en aplicar la segunda parte de lo que hemos visto sobre preprocesamiento:

- **Análisis de datos atípicos,**
- **eliminación de algunas variables predictoras,**
- **transformación de variables**
- **generación de conjunto de entrenamiento y validación)** al conjunto de datos con las 16 variables explicativas que trabajaron en la tarea 1.

Posteriormente hay que ajustar **dos modelos**: uno de regresión Ridge y otro de regresión LASSO, obteniendo las métricas de desempeño.

Observación 1: En este punto NO hacer validación cruzada para obtener el mejor valor de la lambda, se puede tomar igual a 1.

Observación 2: Corrijan lo que comentamos en la revisión de la tarea 1.

Se tiene 7 días para hacer la tarea (a partir del momento en que les envío el correo). Por favor pongan explicaciones de todo lo que hacen, no solamente pongan puro código, de lo contrario no serán acreedores a toda la calificación de esta tarea. Cuiden su redacción y ortografía, vayan acostumbrándose a entregar documentos más “ejecutivos/profesionales”. La tarea se puede hacer directamente en un archivo de google colab, con las explicaciones respectivas en celdas de texto. El colab se tiene que imprimir como pdf.