

Examen 1

Diplomado en Ciencia de Datos — Módulo 4 - Grupo 6

Agosto 2021

Conjunto de datos

El conjunto de datos provisto contiene imágenes sobre casas y sus diferentes habitaciones, así como el precio de la vivienda. Cada casa está representada por cuatro imágenes: dormitorio, baño, cocina y una imagen frontal de la casa.

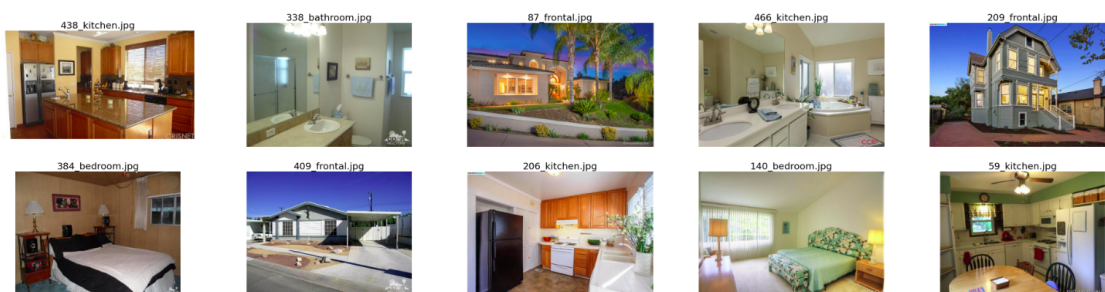


Figura 1: Muestra de imágenes de casas

Además, se proporciona un archivo de texto que contiene los metadatos textuales del conjunto de datos. Cada fila del archivo respeta el número de casa en orden. Los números representan el número de dormitorios, el número de baños, el área de la casa, el código postal y el precio, respectivamente.

	n_rooms	n_bathrooms	area	zipcode	price
0	4	4	4053	85255	869500
1	4	3	3343	36372	865200
2	3	4	3923	85266	889000
3	5	5	4022	85262	910000
4	3	4	4116	85266	971226
...
530	5	2	2066	94531	399900
531	4	3	9536	94531	460000
532	3	2	2014	94531	407000
533	4	3	2312	94531	419000
534	5	3	3796	94531	615000

Figura 2: Metadatos

Actividades

1. Entrene un modelo con Keras para estimar de la mejor forma posible el precio de las casas a partir de las imágenes. Las únicas arquitecturas y capas válidas serán aquellas vistas en clase.
2. Con el mejor modelo entrenado, califique los datos que se encuentran en la carpeta "validate".
3. Responda a las siguientes preguntas:
 - a) ¿Para qué se utiliza *BatchNormalization*?
 - b) ¿Cómo se calculan los valores de la matriz de filtro en la capa convolucional de una CNN?
 - c) ¿Cuál es el objetivo de la capa de Pooling?

Entregables

- (7-10 puntos) Conjunto de validación calificado. Las predicciones deben de encontrarse en formato csv con las columnas: "file_name", "price". La calificación de este rubro dependerá de las calificaciones obtenidas por todo el grupo, es decir, el mayor accuracy obtendrá 10 y el menor obtendrá un 7, el resto se distribuirá en dicho rango de acuerdo con el accuracy obtenido.
- (1.5 puntos) Respuestas a las preguntas en formato txt.