

- ابتدا داده ها را از Kaggle لود کرده و پردازش اولیه انجام میدهیم.
- سپس فیچر هایی که با هدف ما (livingSpace) کورلیشن دارند را به ترتیب صعودی مرتب میکنیم تا فیچر با بیشترین کورلیشن بدست آید.

1. سپس با جستجو روشی برای پیاده سازی k fold پیدا کردم و رگرسیونی برای فیچرها نوشتم به فرمول زیر:

$$\text{livingSpace} = a + b * \text{LivingSpaceRange}$$

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

2. همان روش k fold را پیاده سازی کردم اما برای محاسبه رگرسیون از دستور linearRegression موجود در sklearn استفاده کردم.

3. دو فیچر با کمترین کورلیشن و دو فیچر با بیشترین کورلیشن را از مرحله ی فوق که ذکر شد پیدا کردم و آنها را به همراه فیچر هدف در یک دیتاست جدید ریختم. همان روش k fold را پیاده سازی کردم اما y را تابع هدف گرفتم و x را تمام فیچر های در دیتاست جدید منهای فیچر هدف قرار دادم.

4. مانند سوال 3 اما با فیچرهای متفاوت اما هدف یکسان پیاده سازی کردم.

5. دستوری برای پیاده سازی رگرسیون ریج پیدا کردم و آن را با استفاده از روش k fold ذکر شده پیاده سازی کردم.

6. دستوری برای پیاده سازی رگرسیون لسو پیدا کردم و آن را با استفاده از روش k fold ذکر شده پیاده سازی کردم.