

بعد از خواندن دیتا و انجام مرحله preprocessing که در تمرین قبلی کامل توضیح داده شده بود (شامل : data cleaning, encoding categorical features, reducing memory usage, data type converting, and dropping object features) ۳ ستون فیچر های ذکر شده در سند تمرین (از آنجایی که فیچر heatingType، کتگوریکال می باشد و در مرحله ی preprocessing one-hot encoding بر روی آن اعمال شده خودش شامل ۸ ستون می باشد) را جدا کردیم و به عنوان input variables قرار دادیم و target variable را برابر totalRent قرار دادیم. سپس دیتای ترین و تست را از دیتاست اصلی استخراج کردیم و دیتا ها را توسط standard scaler اسکیل کردیم.

مدل رگرسیون خطی در فایل linreg.py پیاده سازی شده که توضیحات پارامتر ها و متد های آن به طور کامل در فایل مربوطه داده شده، علاوه بر آن نکته ای که در مورد این کلاس وجود دارد این است که این کلاس یک پارامتر اضافی به اسم cost میگیرد که باید از جنس callable باشد و برای مشخص کردن نوع خطا در محاسبه مدل است.

دیتای تولید شده در قسمت بالا بر روی ۳ مدل رگرسیون خطی که پیاده سازی شده بود با خطا های MSE، Absolute error، epsilon sensitive و مدل رگرسیون خطی پکیج sklearn و مدل های Ridge و Lasso پکیج sklearn فیت شدند که نتایج تمامی مدل ها در جدول زیر قابل مشاهده می باشد که تمامی مدل ها تقریباً عملکرد یکسانی بر روی این دیتا داشته اند.

	Model	Train MSE	Test MSE	Train MAE	Test MAE
0	Linear Regression custom MSE	192197.922092	189528.16983	318.174305	318.792085
1	Linear Regression custom AE	192197.922095	189528.17365	318.174295	318.792074
2	Linear Regression custom Epsilon sensitive	192197.922093	189528.150508	318.174284	318.792063
3	LinearRegression	192197.90625	189527.9375	318.185638	318.80365
4	Ridge	192197.9375	189527.828125	318.183441	318.801117
5	Lasso	192205.28125	189672.1875	318.394501	319.03183