Assignment 2: Linear regression and ridge regression

Lab report

Gohil Happy

Roll No: 21IM30006

26th Aug 2023

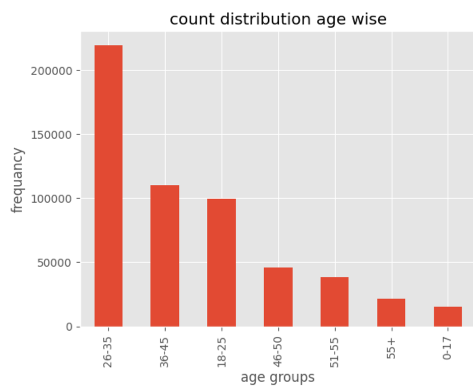
Contents

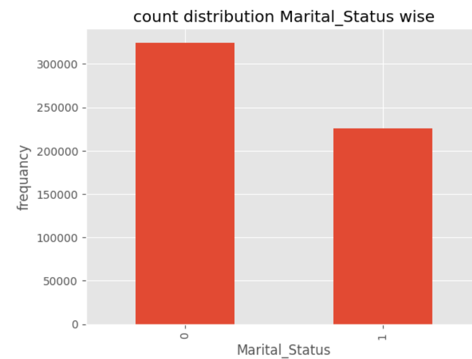
ABSTRACT

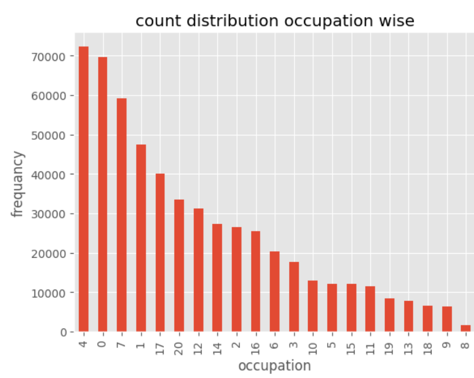
This is report

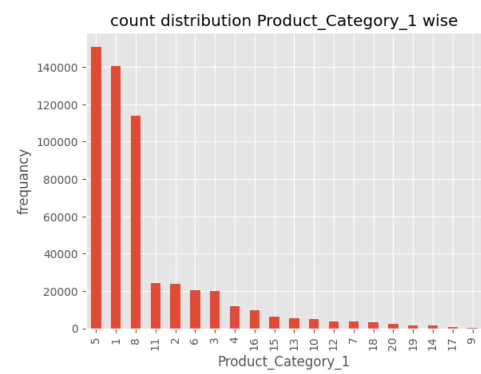
1 Experiment 1: EDA (Exploratory Data Analysis)

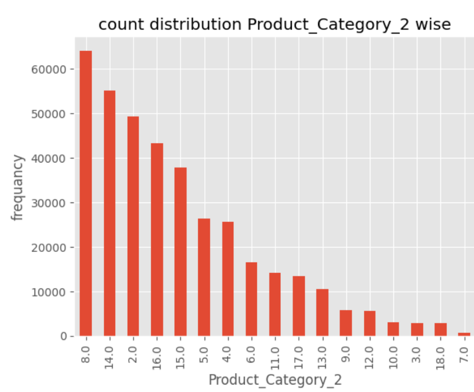
1.1 Plots

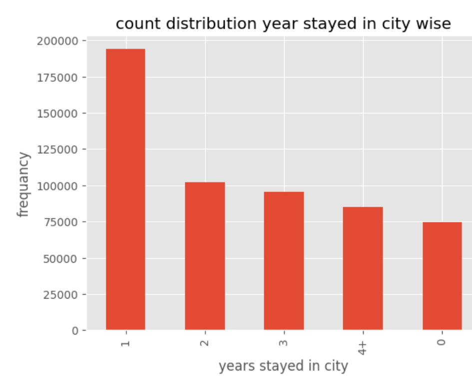


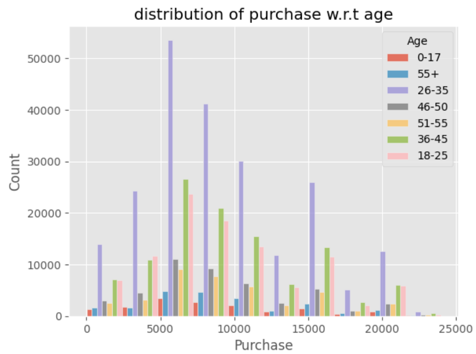


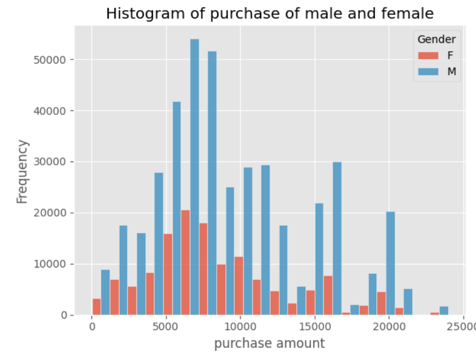


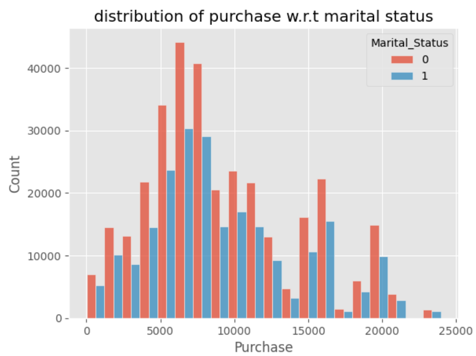


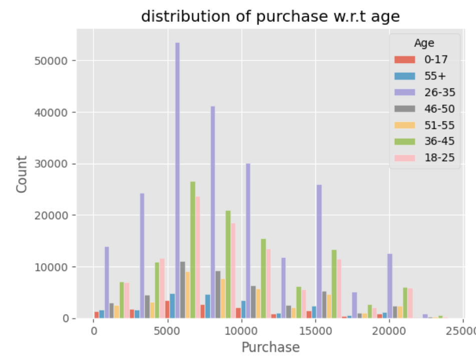






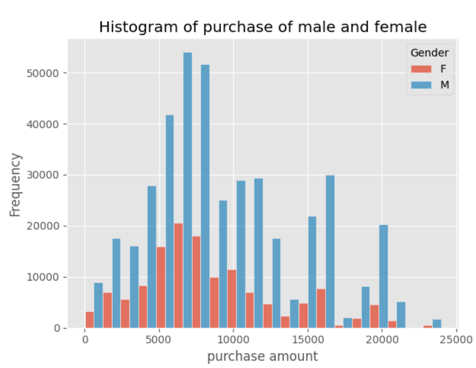


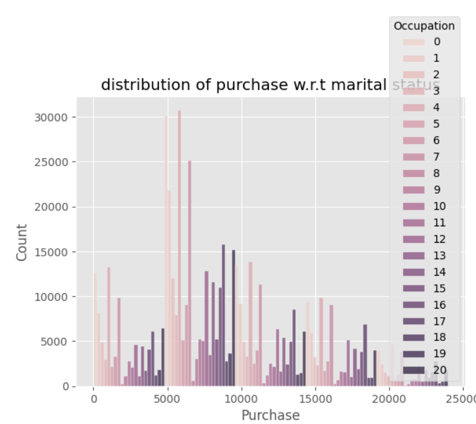




1.2 Procedure to EDA

1. First, dropped the `User_ID` and `"Product_ID"` because IDs do not make any sense in the model.
2. **Handling the NaN values**
   - In `Product_category_3`, we had 69.67% NaN values, so we dropped it.
   - In `Product_category_2`, I used a **forward fill**. From the plot of **"count distribution product_category_2 wise"**, it is evident that using a mode fill would make our model more biased towards one category of product.





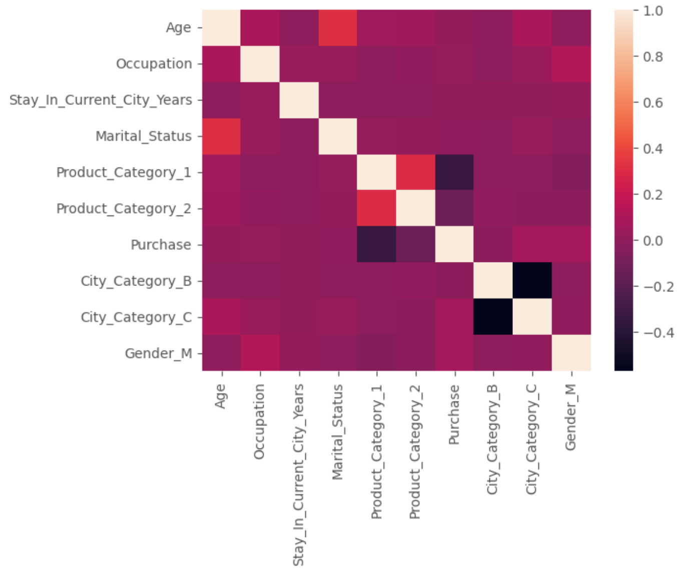


Figure 1: heat map for correlation for diffetent features

	Age	Occupation	Stay_In_Current_City_Years	Marital_Status	Product_Category_1	Product_Category_2	Purchase	City_Category_B	City_Category_C	Gender_M
0	17	10	2	0	3	6.0	8370	0	0	0
1	17	10	2	0	1	6.0	15200	0	0	0
2	17	10	2	0	12	6.0	1422	0	0	0
3	17	10	2	0	12	14.0	1057	0	0	0
4	56	16	4	0	8	14.0	7969	0	1	1
5	35	15	3	0	1	2.0	15227	0	0	1
6	50	7	2	1	1	8.0	19215	1	0	1
7	50	7	2	1	1	15.0	15854	1	0	1
8	50	7	2	1	1	16.0	15686	1	0	1
9	35	20	1	1	8	16.0	7871	0	0	1
10	35	20	1	1	5	11.0	5254	0	0	1
11	35	20	1	1	8	11.0	3957	0	0	1
12	35	20	1	1	8	11.0	6073	0	0	1

Figure 2: Data after EDA

3. Used one-hot encoding for `gender` and `city_category`.
4. Replaced the values for `stay_in_city_years` using the following code: `df['Stay_In_Current_City_Years'].replace('2':2, '4+':4, '3':3, '1':1, '0':0, inplace=True)`
5. For `Age`, replaced using the following code: `df['Age'].replace('0-17':17, '55+':56, '26-35':35, '46-50':50, '51-55':55, '36-45':45, '18-25':25, inplace=True)` becuase this features have **ordinal nature** so there wights matter in model
6. and Occupation, Product category 1 & 2 should remain as it is to avoid feature complexity

2 Experiment 2

- **With out feature scaling**: The value of Mean squared error (MSE) with out feature scaling came out **22081144.710634716**
- **With feature scaling**: The value of Mean squared error (MSE) with out feature scaling came out **22081144.71063472**
- we observe no difference in the value of MSE

3 Experiment 3

Table 1: MSE values for different learning rates for scaled data

sr. no.	lr	MSE
0	0.000 01	$2.884\,403 \times 10^{7}$
1	0.000 10	$2.517\,077 \times 10^{7}$
2	0.001 00	$2.209\,787 \times 10^{7}$
3	**0.01000**	**2.208141e+07**
4	0.100 00	$2.208\,752 \times 10^{7}$
5	0.500 00	$2.215\,061 \times 10^{7}$

Table 2: MSE values for different learning rates for unscaled data

sr. no.	lr	MSE
0	0.00001	2.707129e+07
1	0.00010	2.433942e+07
2	0.00100	2.213294e+07
3	0.01000	NaN
4	0.10000	NaN
5	0.50000	NaN

3.1 Observations

- For **lr = 0.01** the MSE is minimum that means lr = 0.01 is optimal hyper-parameter
- On unscaled data it is observed that for higher value of learning rate algorithm is not converging

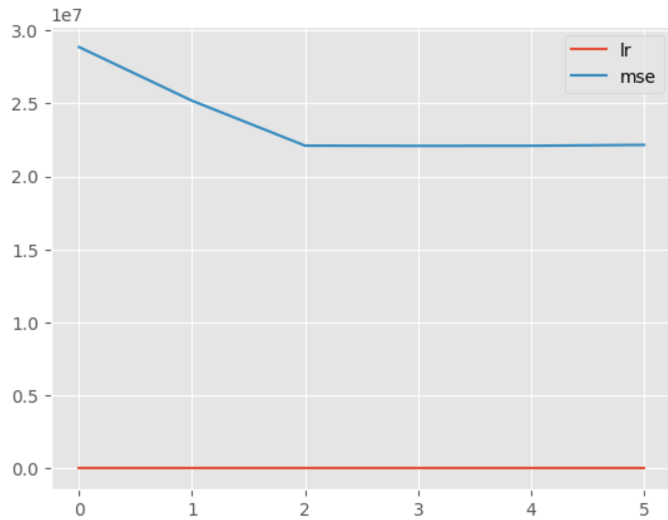


Figure 3: lr vs mse

4 Experiment 4

Applying ridge regression with learning rate = 0.01 on different value of alpha from 0 to 1 with increment of 0.1 we get below table

Table 3: MSE values for different alpha values

Sr. No	Alpha	MSE
0	0.0	$3.165\,392 \times 10^7$
1	0.1	$3.274\,766 \times 10^7$
2	0.2	$3.390\,877 \times 10^7$
3	0.3	$3.512\,085 \times 10^7$
4	0.4	$3.636\,997 \times 10^7$
5	0.5	$3.764\,433 \times 10^7$
6	0.6	$3.893\,397 \times 10^7$
7	0.7	$4.023\,053 \times 10^7$
8	0.8	$4.152\,699 \times 10^7$
9	0.9	$4.281\,752 \times 10^7$
10	1.0	$4.409\,729 \times 10^7$

From plot we can observe the increase in MSE as we increase the value of alpha for ridge regression This happens due to

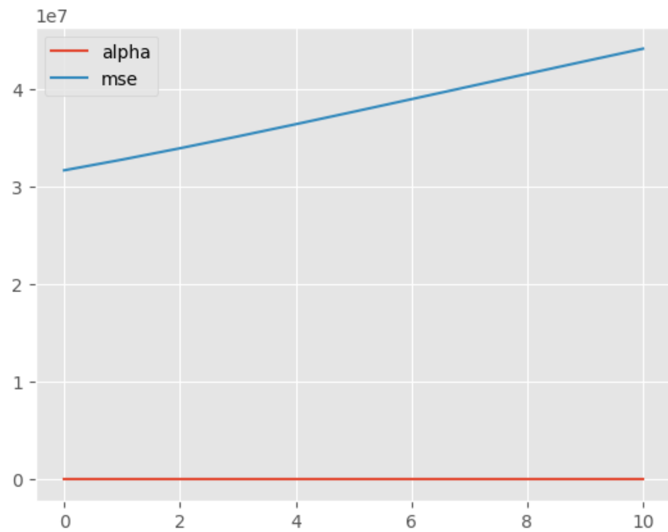


Figure 4: alpha vs mse

5 Experiment 5

Optimal hyper-parameter

$$\text{lr} = 0.01$$

$$\alpha = 0$$

Table 4: MSE values for different models

Sr. No	model	MSE
1	LIN_MODEL_CLOSED	$2.208\,114\,4 \times 10^7$
2	LIN_MODEL_GRAD	$2.208\,141 \times 10^7$
3	LIN_MODEL_RIDGE	$3.165\,392 \times 10^7$

- We observed a increase in MSE as we increase alpha this is due to less number of epochs where wights are updating slowly due to penalization.
- For closed model and gradient decent we observe no difference with respect to grad model