

Project2_2L_NN(Expedia)

April 3, 2022

***** Exploratory Data Analysis *****

----- Displaying head -----

	id	region	latitude	longitude	accommodation_type	cost	\
0	13232	Manhattan	40.71854	-74.00439	Entire home/apt	170	
1	246	Brooklyn	40.64446	-73.95030	Entire home/apt	65	
2	19091	Queens	40.78573	-73.81062	Private room	85	
3	34305	Manhattan	40.73863	-73.98002	Private room	210	
4	444	Manhattan	40.82426	-73.94630	Shared room	75	

	minimum_nights	number_of_reviews	reviews_per_month	owner_id	\
0	5	7	0.56	929983	
1	3	238	2.30	281764	
2	1	0	NaN	19923341	
3	30	0	NaN	200380610	
4	3	38	0.42	745069	

	owned_hotels	yearly_availability
0	1	0
1	1	0
2	1	1
3	65	1
4	3	1

----- Dimensions of dataset -----

(2870, 12)

----- Contents of dataset -----

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2870 entries, 0 to 2869
Data columns (total 12 columns):
#   Column              Non-Null Count  Dtype
#   ...
```

```

---  -----
0   id                2870 non-null   int64
1   region            2870 non-null   object
2   latitude          2870 non-null   float64
3   longitude         2870 non-null   float64
4   accommodation_type 2870 non-null   object
5   cost              2870 non-null   int64
6   minimum_nights    2870 non-null   int64
7   number_of_reviews 2870 non-null   int64
8   reviews_per_month 2194 non-null   float64
9   owner_id          2870 non-null   int64
10  owned_hotels      2870 non-null   int64
11  yearly_availability 2870 non-null   int64

```

dtypes: float64(3), int64(7), object(2)

memory usage: 269.2+ KB

None

----- Summary of Numerical feature -----

	Feature_name	datatype	Count	min	quartile1	Mean	Median	\
0	X	int64	517	1.0	3.0	4.669246	4.00	
1	Y	int64	517	2.0	4.0	4.299807	4.00	
2	FFMC	float64	517	18.7	90.2	90.644681	91.60	
3	DMC	float64	517	1.1	68.6	110.872340	108.30	
4	DC	float64	517	7.9	437.7	547.940039	664.20	
5	ISI	float64	517	0.0	6.5	9.021663	8.40	
6	temp	float64	517	2.2	15.5	18.889168	19.30	
7	RH	int64	517	15.0	33.0	44.288201	42.00	
8	wind	float64	517	0.4	2.7	4.017602	4.00	
9	rain	float64	517	0.0	0.0	0.021663	0.00	
10	area	float64	517	0.0	0.0	12.847292	0.52	

	quartile3	max	Std dev	Skewness	Kurtosis	Range	IQR	\
0	7.00	9.00	2.31	0.04	-1.17	8.00	4.00	
1	5.00	9.00	1.23	0.42	1.42	7.00	1.00	
2	92.90	96.20	5.52	-6.58	67.07	77.50	2.70	
3	142.40	291.30	64.05	0.55	0.20	290.20	73.80	
4	713.90	860.60	248.07	-1.10	-0.25	852.70	276.20	
5	10.80	56.10	4.56	2.54	21.46	56.10	4.30	
6	22.80	33.30	5.81	-0.33	0.14	31.10	7.30	
7	53.00	100.00	16.32	0.86	0.44	85.00	20.00	
8	4.90	9.40	1.79	0.57	0.05	9.00	2.20	
9	0.00	6.40	0.30	19.82	421.30	6.40	0.00	
10	6.57	1090.84	63.66	12.85	194.14	1090.84	6.57	

skewness comment outlier comment

```

0 Fairly symmetric(positive) No outliers
1 Fairly symmetric(positive) Has outliers
2 High negative skewed Has outliers
3 Moderate positive skewed Has outliers
4 High negative skewed Has outliers
5 High positive skewed Has outliers
6 Fairly symmetric(negative) Has outliers
7 Moderate positive skewed Has outliers
8 Moderate positive skewed Has outliers
9 High positive skewed Has outliers
10 High positive skewed Has outliers

```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 2870 entries, 0 to 2869
```

```
Data columns (total 10 columns):
```

#	Column	Non-Null Count	Dtype
0	region	2870 non-null	object
1	latitude	2870 non-null	float64
2	longitude	2870 non-null	float64
3	accommodation_type	2870 non-null	object
4	cost	2870 non-null	int64
5	minimum_nights	2870 non-null	int64
6	number_of_reviews	2870 non-null	int64
7	reviews_per_month	2194 non-null	float64
8	owned_hotels	2870 non-null	int64
9	yearly_availability	2870 non-null	int64

```
dtypes: float64(3), int64(5), object(2)
```

```
memory usage: 224.3+ KB
```

```
----- Summary of Numerical feature -----
```

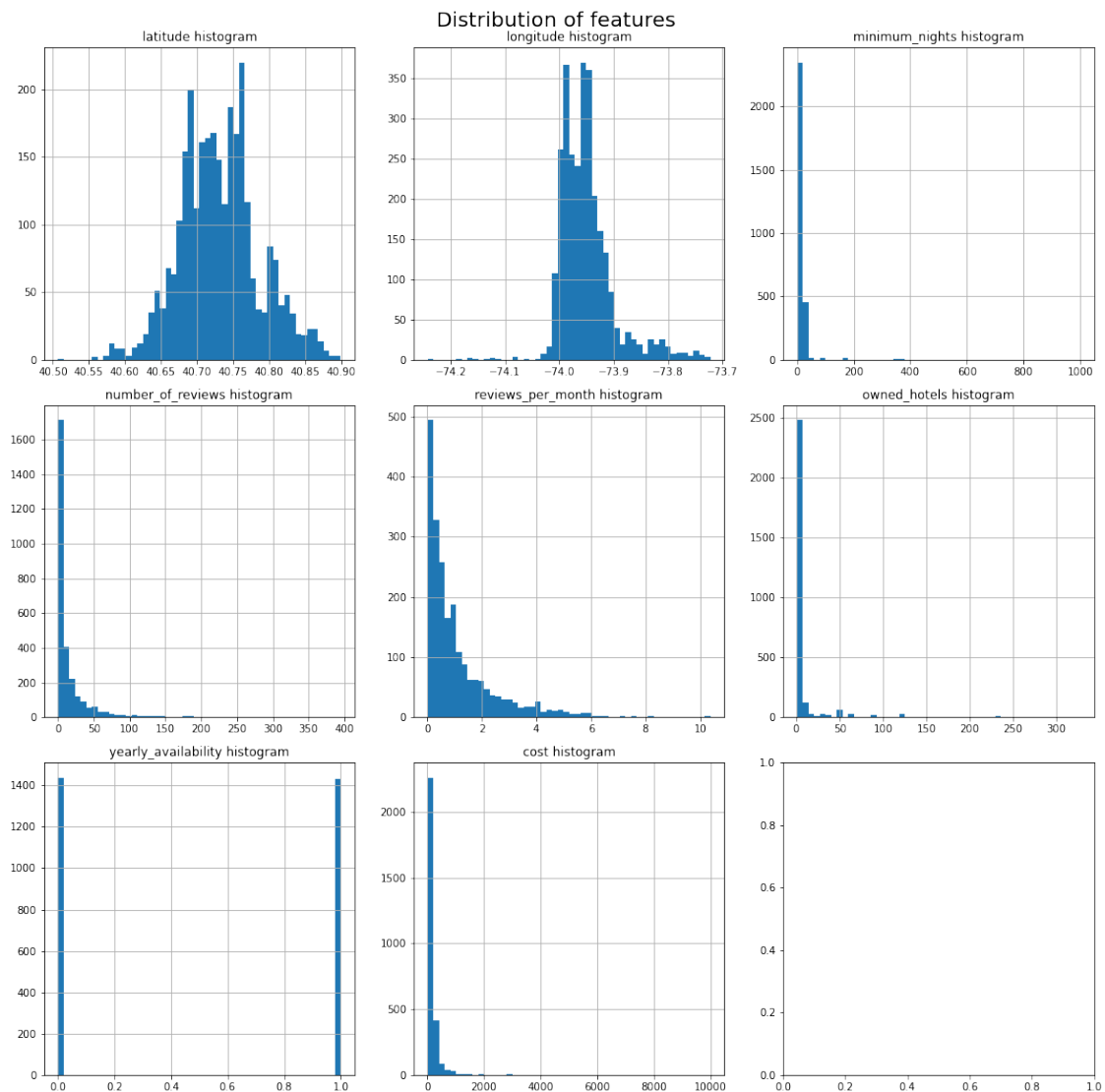
	Feature_name	datatype	Count	min	quartile1	Mean	\
0	latitude	float64	2870	40.50708	40.692463	40.731224	
1	longitude	float64	2870	-74.24285	-73.984003	-73.950158	
2	cost	int64	2870	10.00000	75.000000	195.943206	
3	minimum_nights	int64	2870	1.00000	1.000000	11.530314	
4	number_of_reviews	int64	2870	0.00000	1.000000	16.315331	
5	reviews_per_month	float64	2194	0.01000	0.240000	1.157502	
6	owned_hotels	int64	2870	1.00000	1.000000	8.411498	
7	yearly_availability	int64	2870	0.00000	0.000000	0.498606	

	Median	quartile3	max	Std dev	Skewness	Kurtosis	Range	\
0	40.72825	40.762658	40.89873	0.05	0.17	0.21	0.39165	
1	-73.95672	-73.934202	-73.72173	0.05	1.36	4.43	0.52112	
2	120.00000	200.000000	9999.00000	406.18	13.01	232.35	9989.00000	
3	3.00000	6.000000	999.00000	37.97	11.87	210.77	998.00000	

4	4.00000	16.00000	395.00000	32.48	4.27	25.44	395.00000
5	0.65000	1.53000	10.37000	1.36	2.16	5.81	10.36000
6	1.00000	3.00000	327.00000	27.11	6.95	62.60	326.00000
7	0.00000	1.00000	1.00000	0.50	0.01	-2.00	1.00000

	IQR	skewness	comment	outlier	comment
0	0.070195	Fairly	symmetric(positive)	Has	outliers
1	0.049800	High	positive skewed	Has	outliers
2	125.000000	High	positive skewed	Has	outliers
3	5.000000	High	positive skewed	Has	outliers
4	15.000000	High	positive skewed	Has	outliers
5	1.290000	High	positive skewed	Has	outliers
6	2.000000	High	positive skewed	Has	outliers
7	1.000000	Fairly	symmetric(positive)	No	outliers

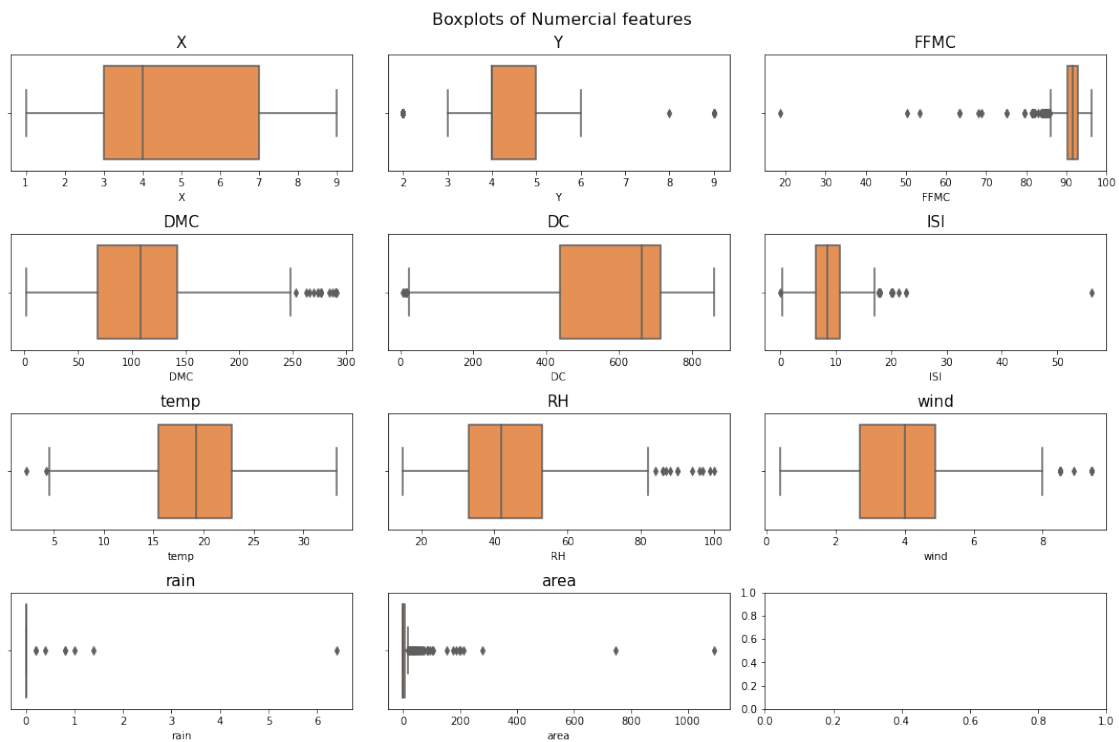
None



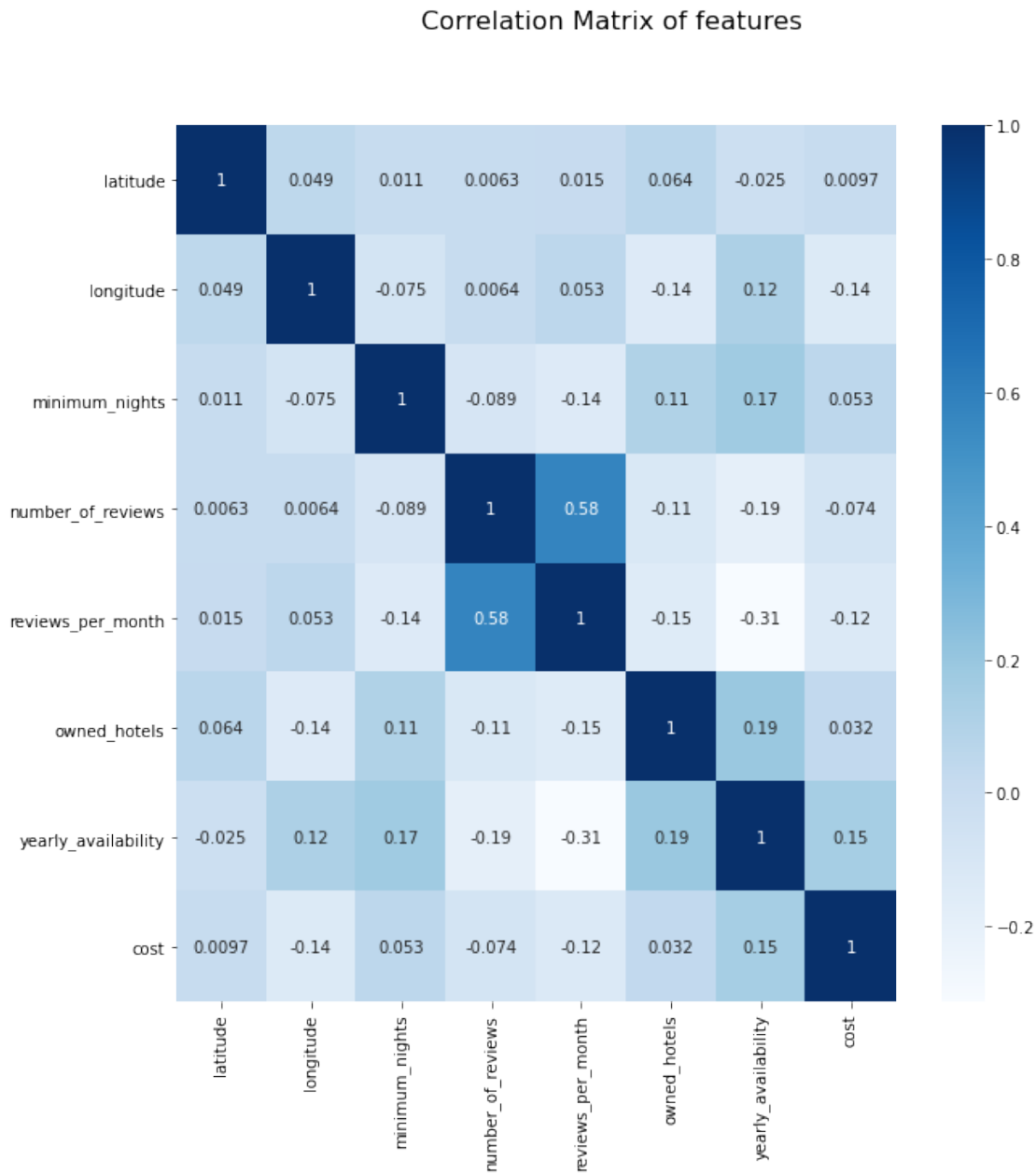
```
[8]: latitude          0
longitude            0
minimum_nights      0
number_of_reviews   0
reviews_per_month   676
owned_hotels        0
yearly_availability  0
cost                0
dtype: int64
```

```
[9]: latitude          0
longitude            0
minimum_nights      0
number_of_reviews   0
reviews_per_month   0
owned_hotels        0
yearly_availability  0
cost                0
dtype: int64
```

None



[10]: <AxesSubplot:>



```
[12]: latitude longitude minimum_nights number_of_reviews \
0 40.71854 -74.00439 5 7
1 40.64446 -73.95030 3 238
2 40.78573 -73.81062 1 0
3 40.73863 -73.98002 30 0
4 40.82426 -73.94630 3 38
... ..
```

2865	40.74316	-73.98038	2	0
2866	40.73523	-73.99465	3	2
2867	40.76619	-73.98987	3	17
2868	40.74637	-73.97207	30	0
2869	40.79208	-73.96482	30	24

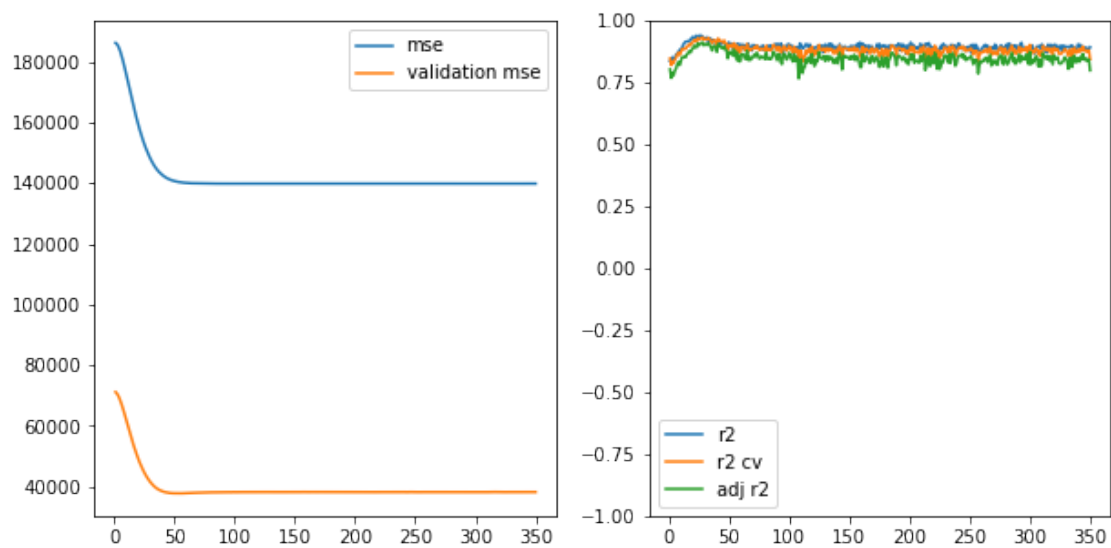
	reviews_per_month	owned_hotels	yearly_availability	cost
0	0.56	1	0	170
1	2.30	1	0	65
2	0.00	1	1	85
3	0.00	65	1	210
4	0.42	3	1	75
...
2865	0.00	1	1	400
2866	0.07	1	1	180
2867	0.67	1	0	179
2868	0.00	49	1	200
2869	0.33	11	1	1000

[2870 rows x 8 columns]

```
[24]: ['latitude',
       'longitude',
       'number_of_reviews',
       'reviews_per_month',
       'yearly_availability']
```

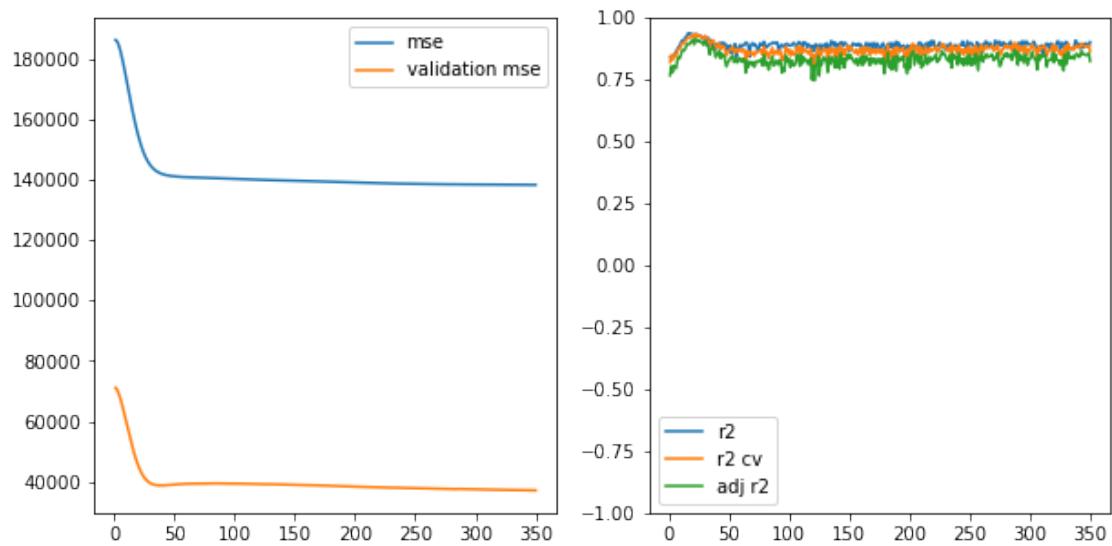
Building Neural nets with linear activation function

2L Neural net with linear activation function



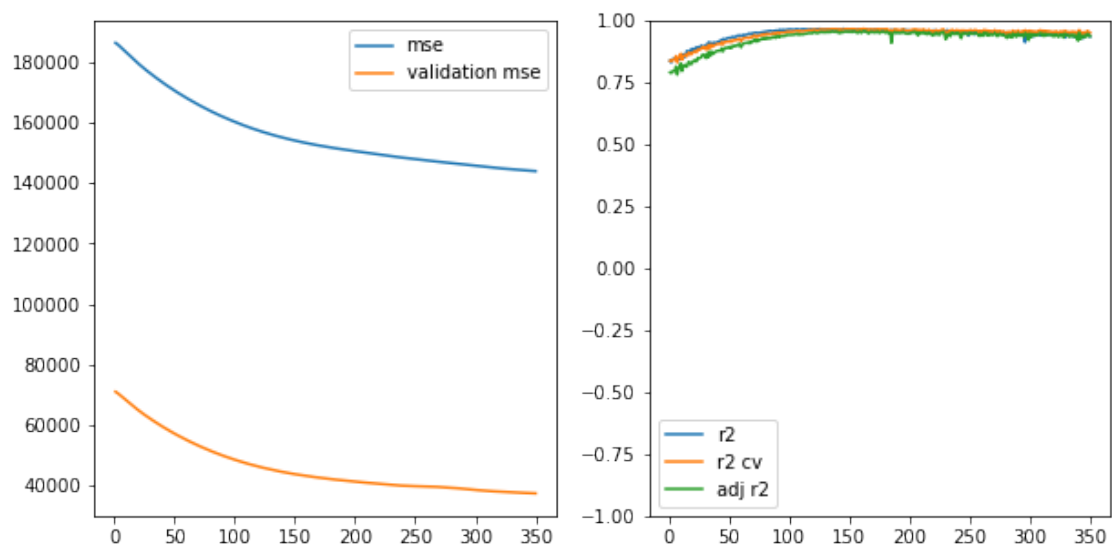
Building Neural nets with relu activation function

2L Neural net with relu activation function



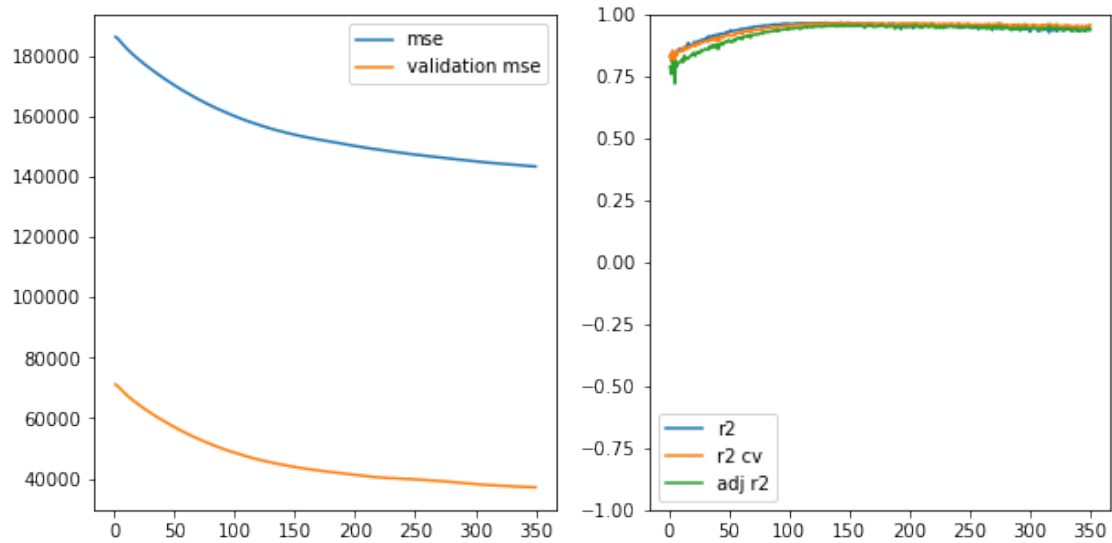
Building Neural nets with sigmoid activation function

2L Neural net with sigmoid activation function



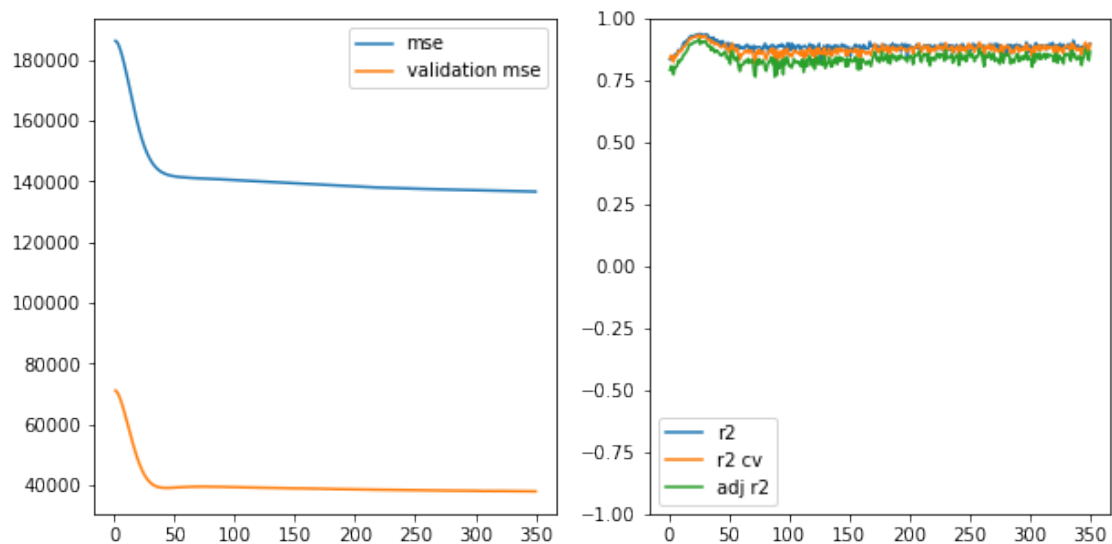
Building Neural nets with tanh activation function

2L Neural net with tanh activation function



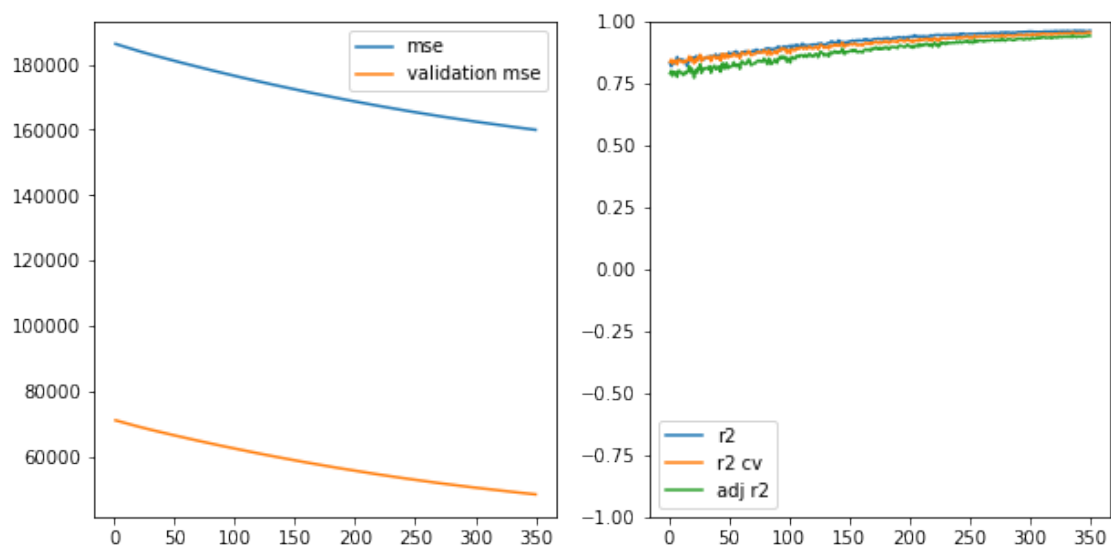
Building Neural nets with elu activation function

2L Neural net with elu activation function



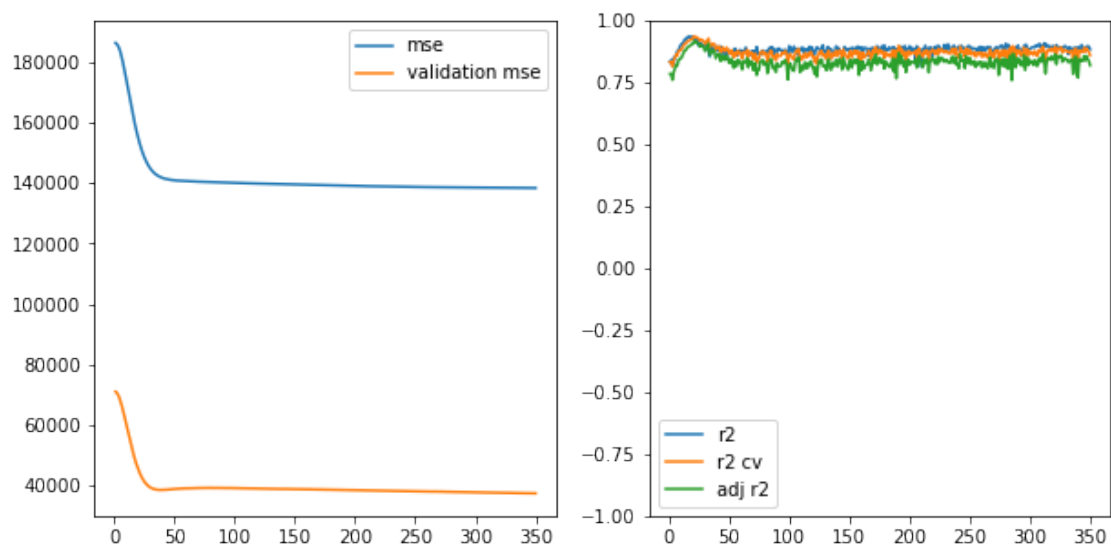
Building Neural nets with softmax activation function

2L Neural net with softmax activation function



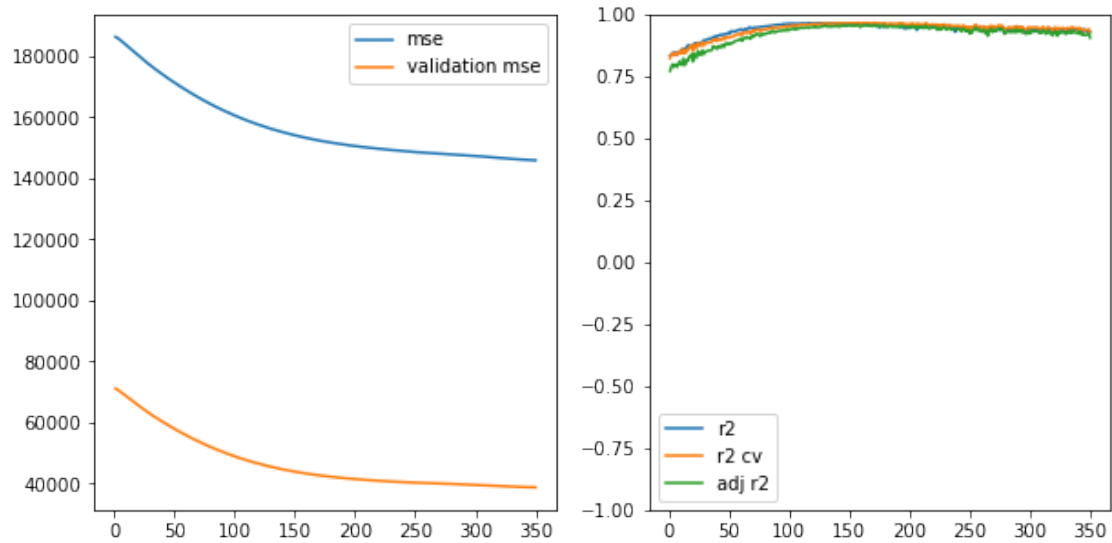
Building Neural nets with softplus activation function

2L Neural net with softplus activation function



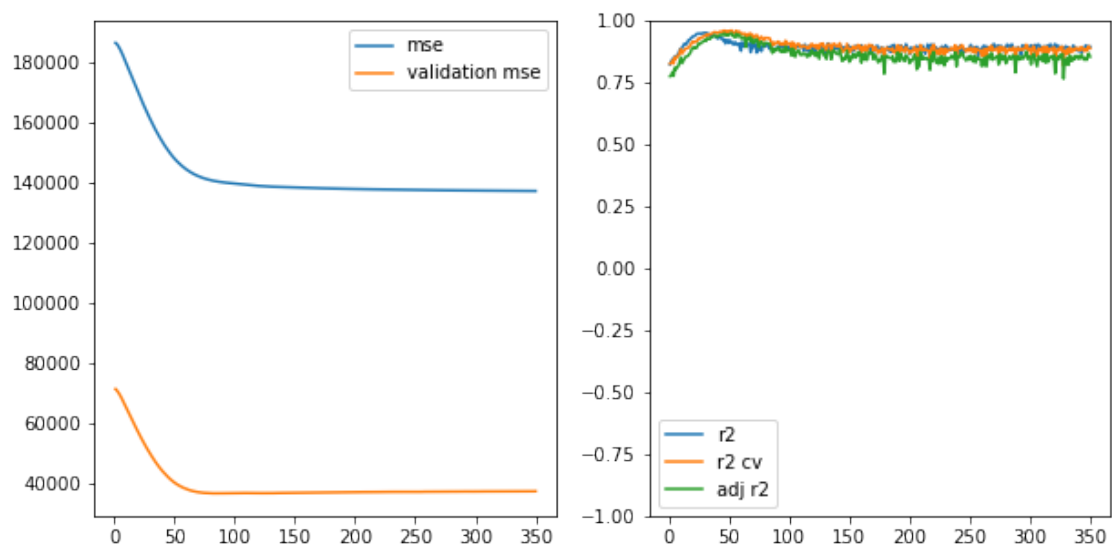
Building Neural nets with softsign activation function

2L Neural net with softsign activation function



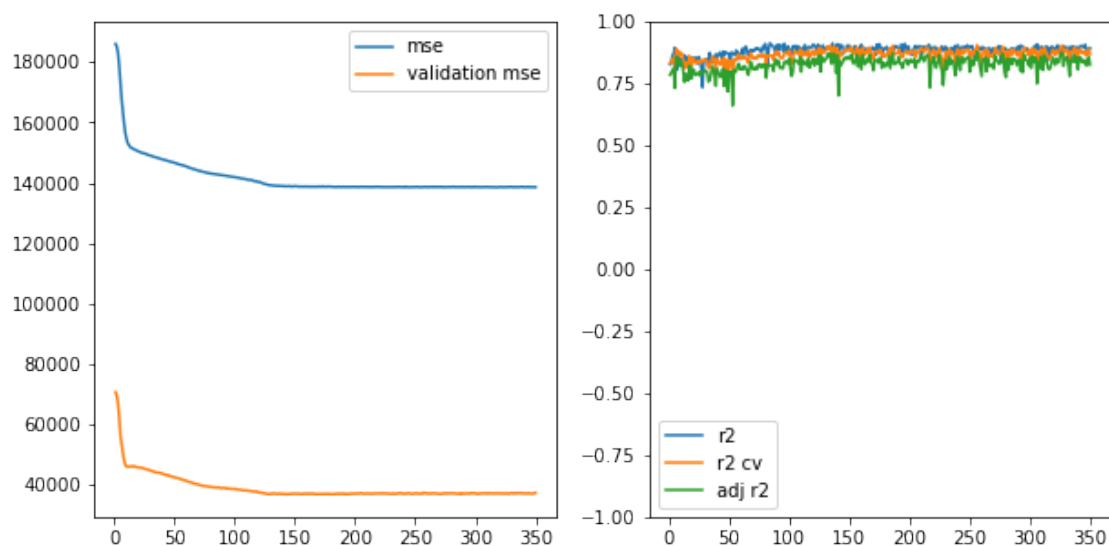
Building Neural nets with selu activation function

2L Neural net with selu activation function



Building Neural nets with exponential activation function

2L Neural net with exponential activation function



***** R2 scores of various activation functions

R2 for NN using linear activation is 4.0721444967848015

Adj R2 for NN using linear activation is 3.880576138316083

R2 CV for NN using linear activation is 4.164673375426742

R2 for NN using relu activation is 4.533901551369035

Adj R2 for NN using relu activation is 4.343255323813111

R2 CV for NN using relu activation is 4.617202634190953

R2 for NN using sigmoid activation is -4.093286648161243

Adj R2 for NN using sigmoid activation is -4.301161409315846
R2 CV for NN using sigmoid activation is -5.778938288477176

R2 for NN using tanh activation is -4.055570007207532
Adj R2 for NN using tanh activation is -4.263369448060672
R2 CV for NN using tanh activation is -5.7009987381414495

R2 for NN using elu activation is 4.6970583972679725
Adj R2 for NN using elu activation is 4.50673799466641
R2 CV for NN using elu activation is 4.6502039977132315

R2 for NN using softmax activation is -11.84608947996255
Adj R2 for NN using softmax activation is -12.069446623207615
R2 CV for NN using softmax activation is -12.823409262176755

R2 for NN using softplus activation is 4.637647438055426
Adj R2 for NN using softplus activation is 4.447208391501367
R2 CV for NN using softplus activation is 4.762533690421522

R2 for NN using softsign activation is -4.2228842981810955
Adj R2 for NN using softsign activation is -4.431017866425102
R2 CV for NN using softsign activation is -5.965589577803154

R2 for NN using selu activation is 4.813977022133309
Adj R2 for NN using selu activation is 4.6238901065509435
R2 CV for NN using selu activation is 4.806151571338346

R2 for NN using exponential activation is 5.433883362665082
Adj R2 for NN using exponential activation is 5.2450344028301625
R2 CV for NN using exponential activation is 5.047698193292993

[NbConvertApp] Converting notebook Project2_2L_NN(Expedia).ipynb to pdf
[NbConvertApp] Writing 380766 bytes to Project2_2L_NN(Expedia).pdf