FML Assignment

Linear Regression and Gradient Descent

Team Members:

NAME	ROLL NUMBER
Sagar Poudel	213051001
Arun Babu PT	213050059
Vishal Kumar	213050081
Kumar Shaswat	213050013
Tanmay Joshi	213050018

Loss Obtained by our ML Model on the Development set:

1. Losses with Analytical Solution:

   train loss: 7737.621323901388,

   dev_loss: 10691.868733133477

2. Losses with Gradient Descent Solution:

   train loss: 9378.323271265144,

   dev_loss: 12165.69671792003

Stopping Criteria:

Gradient descent stopping criteria used:

(lr=0.25, C=0, batch_size=32, max_steps=170000, eval_steps=1000)

1. MSE Loss with Early Stopping:

   dev_loss: 12165.69671792003

   train loss: 9378.323271265144,

2. MSE Loss without Early Stopping:

   dev_loss: 12596.160135865404

   train _loss: 9488.937041704587

Checking if the new weights are giving less weight than the previous weight.

$$W_{new} = MSE_{W_{old}} > MSE_{W_{new}} ? W_{new} : W_{old}$$

When MSE old is not changed for more than 500 iterations then, the loop will stop and use the best gradient solution.

Basis Functions:

Basis function is used for the feature which has higher correlation with target value. New features have been derived using the given features.

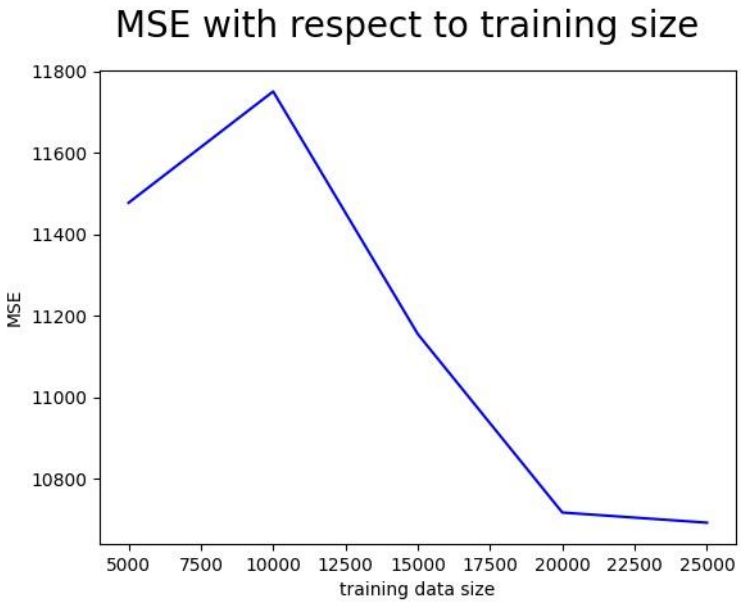
With only 2 Features combinations:

train loss: 9490.81201906974, **dev_loss**: 16629.757529553026

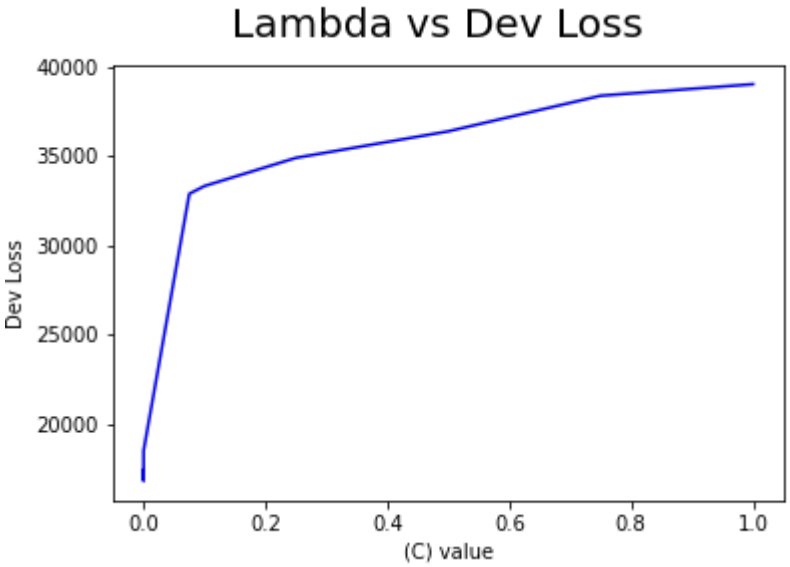
With 3 features combinations:

train loss: 7744.936827234496, **dev_loss**: 10691.868733133477

Plot of MSE on Dev-Set for different values of C (including C=0)



Plot of MSE loss (Y-axis) with various sizes of Training set (X-axis)



Feature Importance:

Out of the 14 available features, we have considered the following features for most important for our ML model: -

- Brightness - VIIRS I-4 channel brightness temperature of the fire pixel measured in Kelvin. This factor is selected because brightness of fire significantly affects its detection by satellites.
- Track - Scan and track reflect actual pixel size and hence was important factor for consideration.
- Confidence - It is important to our model as data points with higher confidence will obviously be more conclusive than data points with low confidence.
- Bright_t31 - I-5 Channel brightness temperature of the fire pixel measured in Kelvin. Higher the temperature in Kelvin, more is the intensity of fire and hence more is its frp.

We have used other features after doing some feature engineering.

- Daynight: More preference to day
- Satellite: preference is given to Aqua than Terra.

Least important feature:

1. Instrument

Feature importance is taken on the basis of correlation of feature with target value.

Have used **df.corr()** to fetch the correlation of each feature with target value

	Unnamed: 0	latitude	longitude	brightness	scan	track
Unnamed: 0	1.000000	-0.001042	0.011708	-0.000918	-0.001485	-0.001631
latitude	-0.001042	1.000000	-0.423952	-0.091418	0.090152	0.098467
longitude	0.011708	-0.423952	1.000000	-0.110053	-0.137715	-0.144127
brightness	-0.000918	-0.091418	-0.110053	1.000000	-0.100590	-0.107098
scan	-0.001485	0.090152	-0.137715	-0.100590	1.000000	0.983342
track	-0.001631	0.098467	-0.144127	-0.107098	0.983342	1.000000
acq_time	-0.002693	-0.240641	0.191910	-0.325280	-0.025701	-0.021958
confidence	0.001579	-0.066814	-0.063306	0.524774	-0.093754	-0.100390
bright_t31	-0.001171	0.192400	-0.315480	0.684836	-0.091400	-0.092033
frp	-0.006257	-0.095640	-0.064630	0.638042	0.186877	0.181642

	acq_time	confidence	bright_t31	frp
Unnamed: 0	-0.002693	0.001579	-0.001171	-0.006257
latitude	-0.240641	-0.066814	0.192400	-0.095640
longitude	0.191910	-0.063306	-0.315480	-0.064630
brightness	-0.325280	0.524774	0.684836	0.638042
scan	-0.025701	-0.093754	-0.091400	0.186877
track	-0.021958	-0.100390	-0.092033	0.181642
acq_time	1.000000	-0.064113	-0.466239	-0.085942
confidence	-0.064113	1.000000	0.260554	0.249365
bright_t31	-0.466239	0.260554	1.000000	0.386822
frp	-0.085942	0.249365	0.386822	1.000000

Climb the leader board:

Key Observation:

We have observed that if the learning rate is high, the MSE will converge very quickly at the beginning and remain constant or might diverge in the opposite direction. If using the less learning rate, it is taking time to converge.

References:

1. https://www.analyticsvidhya.com/blog/2020/12/feature-engineering-using-pandas-for-beginners/
2. https://machinelearningmastery.com/standardscaler-and-minmaxscaler-transforms-in-python/
3. https://www.geeksforgeeks.org/pandas-tutorial/
4. https://www.geeksforgeeks.org/numpy-tutorial/
5. https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.to_csv.html