CS419(M): Programming Assignment-1

Prateek Garg, 20D070060

31st March 2022

1 Preleminary Questions

- Since we are predicting a single quantity, the quality of wine based on given data, this must treated as a regression problem only with some modified output values. CLassification deals with categorizing data points into seperate classes which is not the case here
- Another metric can be **Mean Absolute error**,

$$\frac{\sum_{i=1}^{n} |y_i - H(x_i)|}{n}$$

  This can be interpreted as average of distances from the true value of quality.

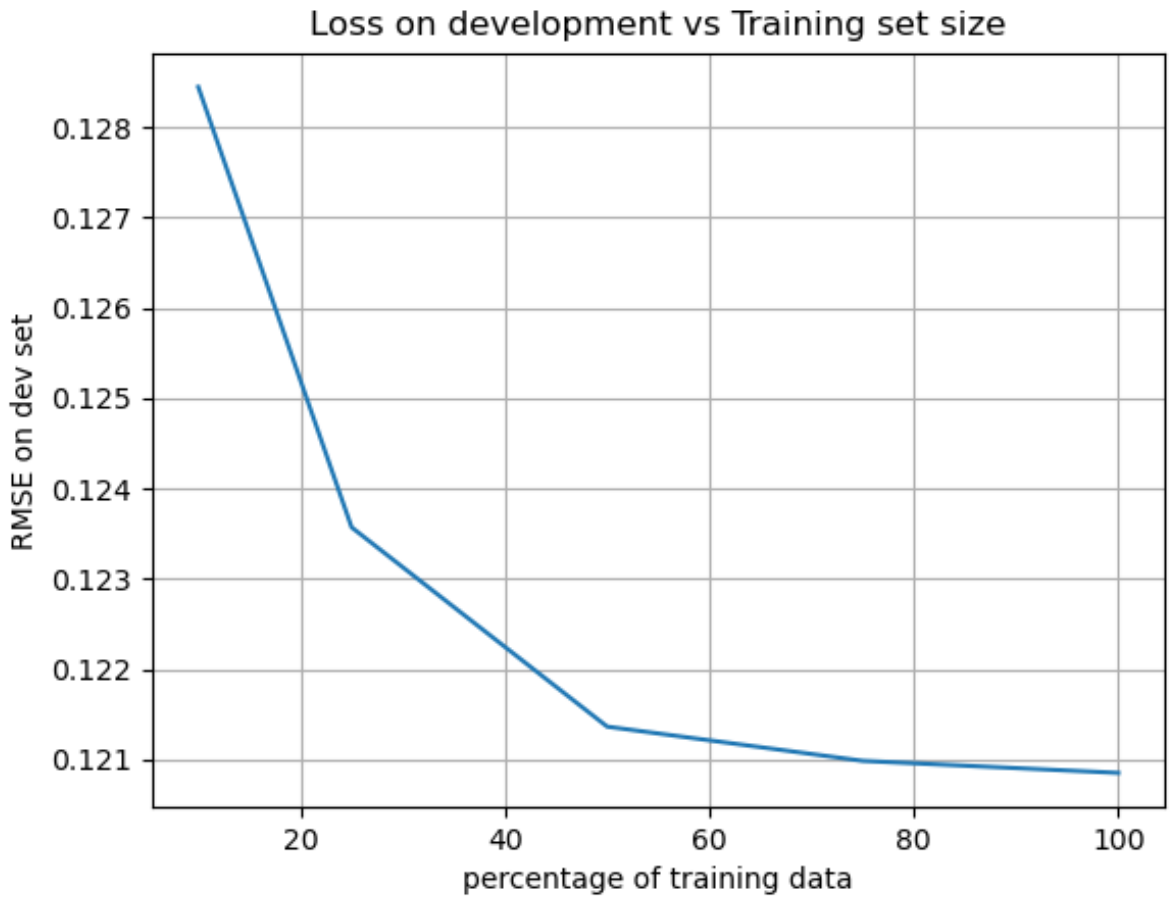
2 Gradient Descent

- Root Mean Squared error on dev test: 0.1207951
- Absolute difference between the following two calls, `compute_RMSE(phi,w1, y)` and `compute_RMSE(phi, w2, y)`:1.15264436e-05
- We used `grad_norm_threshold=0.0001` which is upper limit of gradient of loss with respect to **w**. As soon as gradient goes below this value we stop the process
- Absolute difference between the following two calls, `compute_RMSE(phi,w2, y)` and `compute_RMSE(phi, w3, y)`:1.8945e-04 with learning rate=0.005 and number of iterations=50000

3 Gradient Descent with p-norm regularization

- Root Mean Squared error on dev test with **p=2**: 0.12809152
- Root Mean Squared error on dev test with **p=4**: 0.12197416

4 Training Data size vs RMSE



5 Features

From closed form solution, we get
`w*=[0.0758 0.1375 -0.3049 0.0416 0.3566 -0.0410`
`0.1659 -0.0662 -0.4573 0.1109 0.1342 0.2480]`
with `b*= 0.4079`
Since the features are normalised the weights given by linear regression directly corresponds to the effect of a particular feature on the value.

- The two most useful features thus are: `pH, acidity`
- The two least useful features thus are: `citric acid, chlorides`

6 Regression using open-source library implementations

- We have used `SGDRegressor` from SK-learn library.
- Root Mean Squared error on dev test: 0.122121