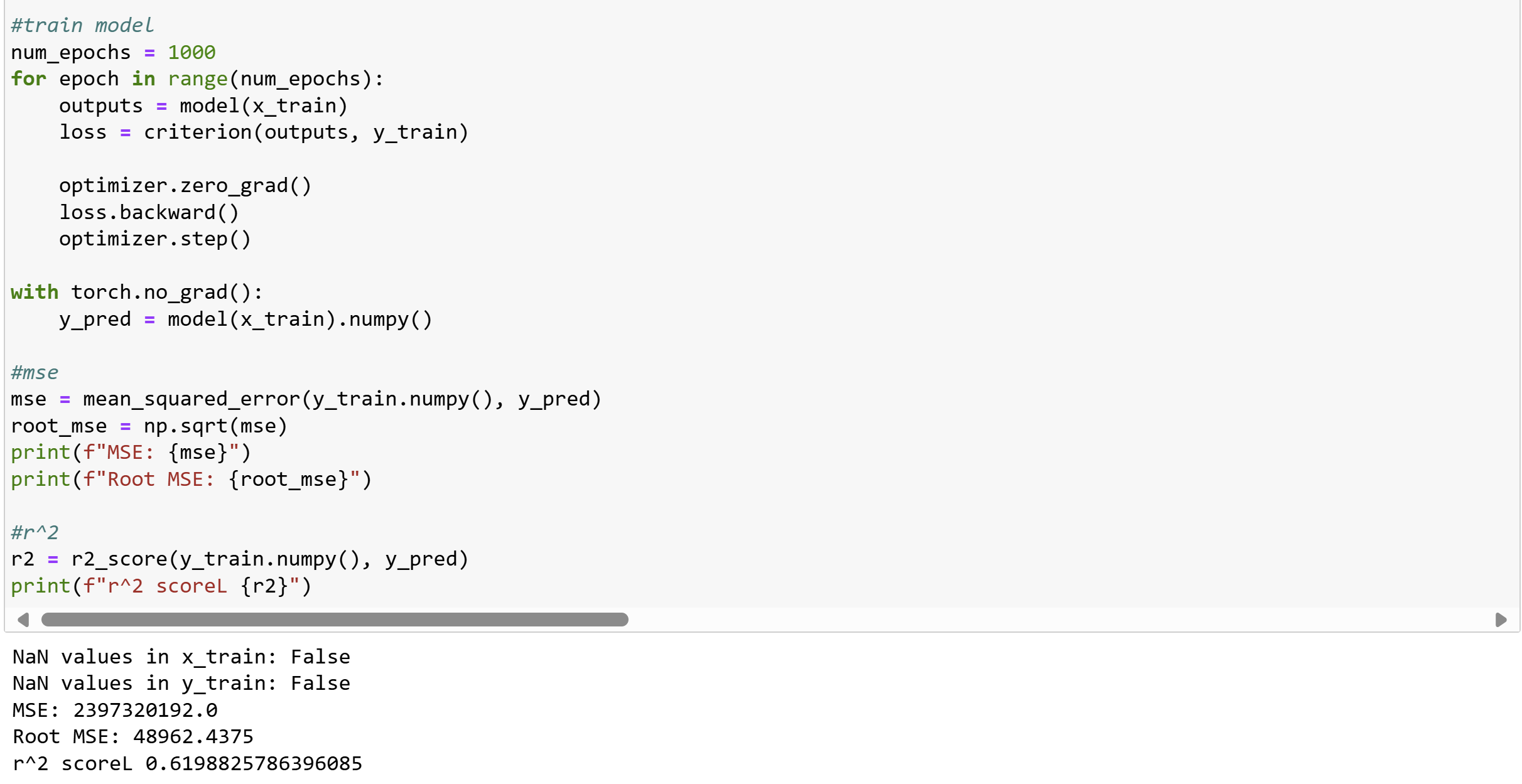
Homework 2 – Report

**Programming Exercises**

**Continuation of House Prices from Hw1**

1a.





1b. The OLS report did considerably worse than sklearn. The root MSE was 48962 while in sklearn the root MSE was around 23417. Gradient descent is more applicable in linear regression when there’s large data sets and more complex loss functions to worry about.

**Data generating distribution and convergence of linear regression**

2a. A screenshot of a computer program

Description automatically generated

To generate synthetic data, I made variables with all the means, std devs., alpha and beta. Then I filled out X and epsilon with the library using the mean and std. dev. since these are not static variables like alpha and beta. Then I made a variable Y, to hold the function. I used a linear regression model to predict the Y values and put that in y\_predict. I then made this into a function that returned the r2 and coefficients of the model so I could use it easily in part b.

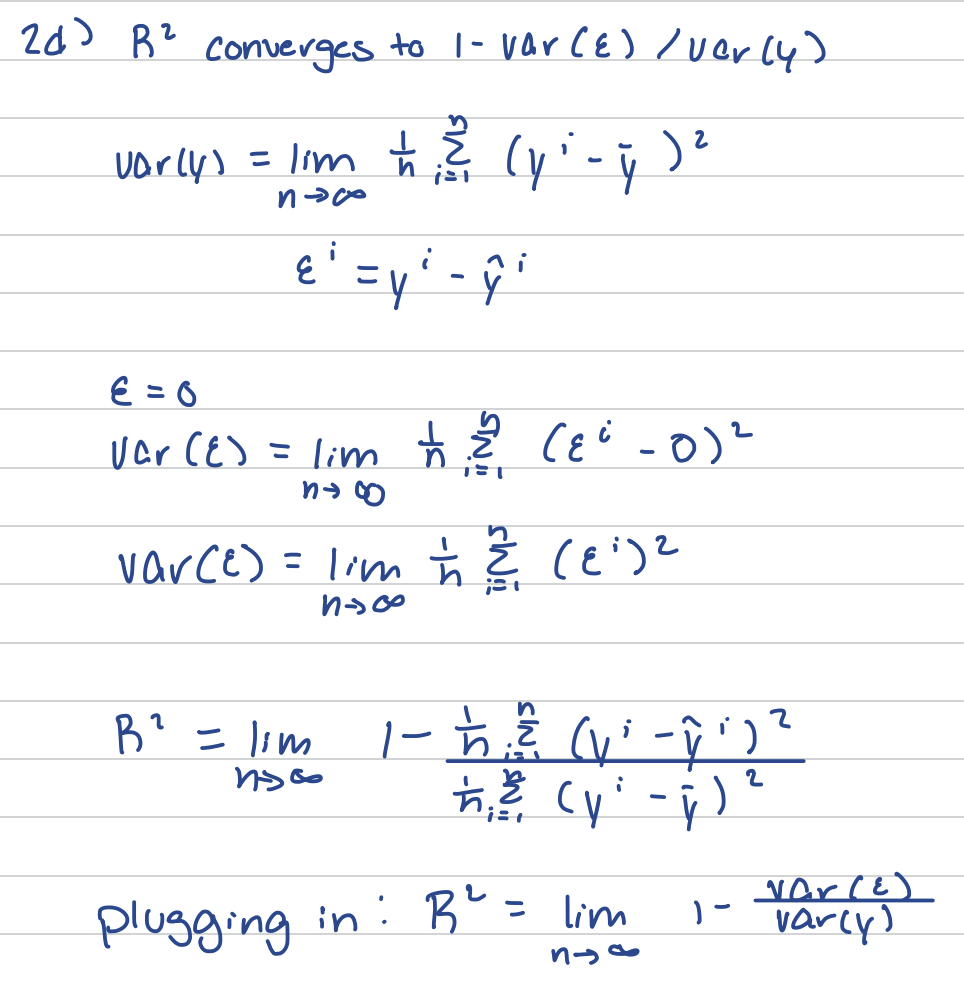
2b.

A screenshot of a computer code

Description automatically generated

2c. From part 2b, I observe that beta seems to converge to 0.499 and alpha converges close to 20.07. However, r2 does not converge to 0 but rather 0.35. This means that the model does not fit the data very well.

2d.



2e. Extra Credit

2f. No I don’t think a better model exists. The relationship between X and Y is linear (assumed) and adding any more parameters would only add noise rather than making the r2 better.

2g. Some characteristics would be having data that truly fits a linear model, not having a lot of noise, and independence between the variables.

**Binary Classification on Text Data**

3a. There are 7613 training and 3263 test points. The percentage of training tweets that are real disasters is 42.97% and not real disasters is 57.03%

A close-up of numbers

Description automatically generatedA close-up of a number

Description automatically generated

A screenshot of a computer code

Description automatically generated

3b.

A screenshot of a computer code

Description automatically generated

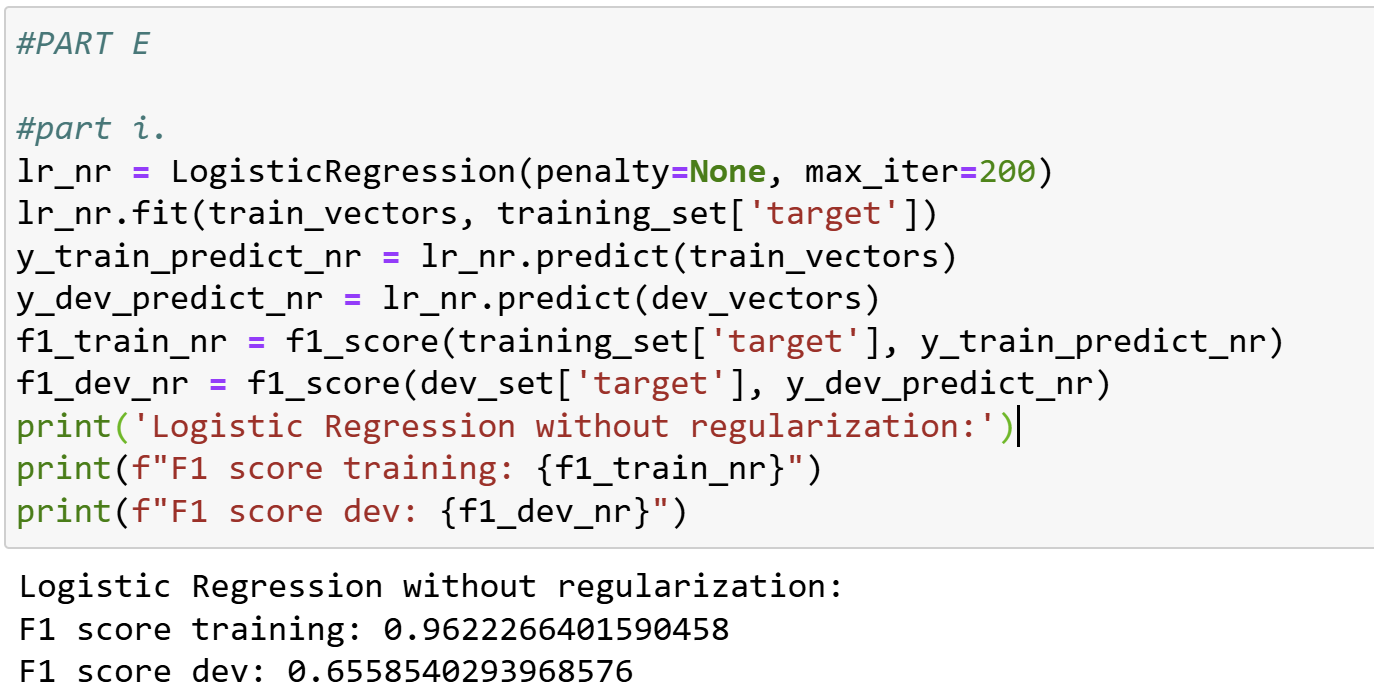
3c. I have decided to also remove # and hashtags in addition to @ and urls and made a function to clean the data ad run it through the training set and dev set

A screenshot of a computer code

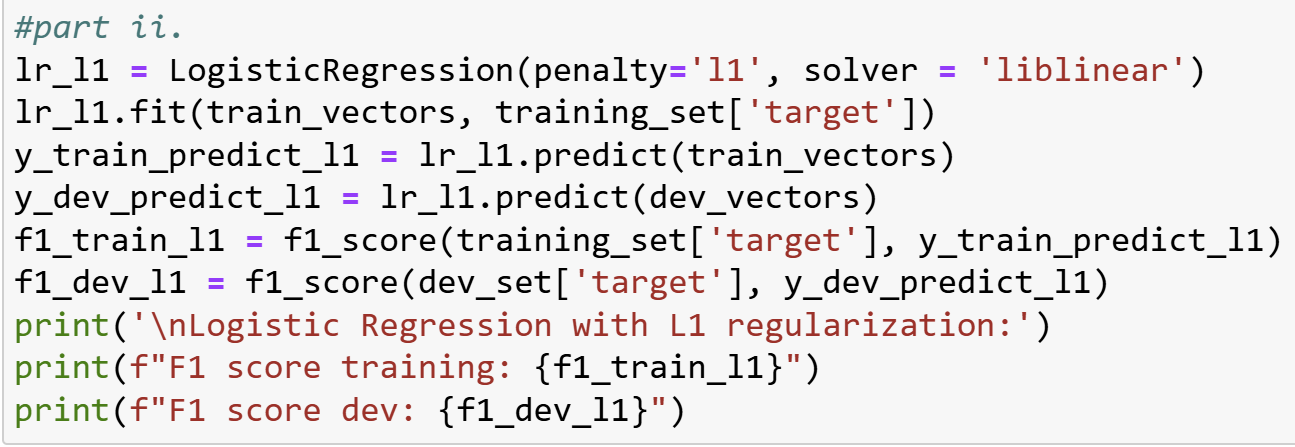
Description automatically generated3d. A screenshot of a computer code

Description automatically generated

3e.



I observe overfitting because the training score is high but dev score is low which means that it learned the training very well to the point it cant perform in dev set.



A black text on a white background

Description automatically generated

Since both F1 scores are similar, this means the model is performing well in training and developing.

A computer code with red and black text

Description automatically generated

A close-up of numbers

Description automatically generated

Since both F1 scores are similar, this means the model is performing well in training and developing.

1. The one that performed best on training was the logistic regression model without regularization terms (f1 = 0.96). The one that performed best on the dev set was regression model with L1 regularization (f1 = 0.73). I observed overfitting with the logistic regression without regularization and regularization did help. Before the difference between the f1 scores of the training and dev set was 0.31 and after the difference between the f1 scores once it was regularized was closer to 0.1 difference.

A screenshot of a computer code

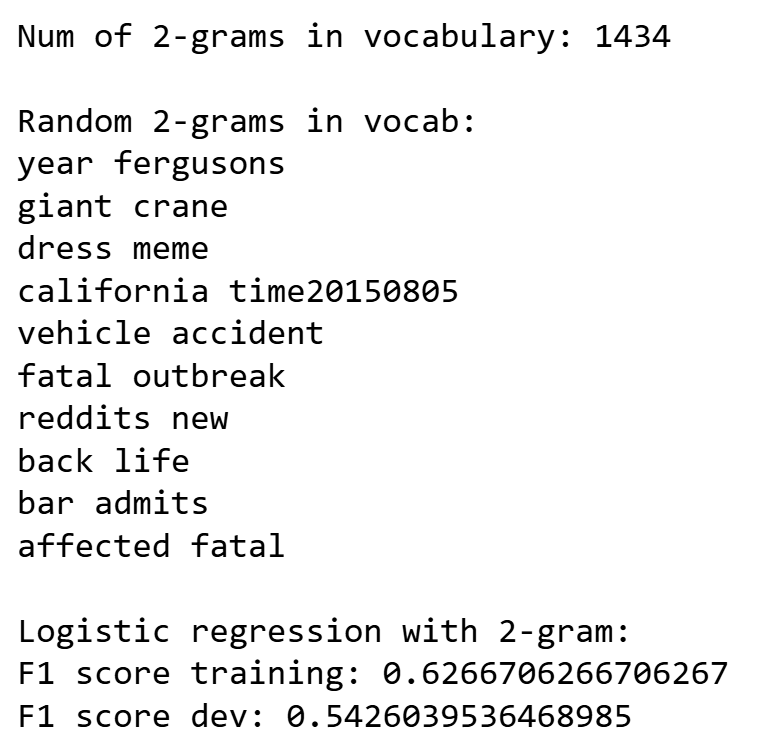
Description automatically generated

A screenshot of a computer

Description automatically generated

The most important to determine whether there’s a disaster are: typhoon, wildfire, Hiroshima, spill, and migrant

3f. 



I chose the threshold M as 3 because it’s not a very large dataset and I feel like 2 times could be a coincidence of people writing the same words but a third person verifies whether it is a real accident or not. This model does worse than bag of models because the F1 score of l1 dev, which was the best model was 0.73 and the F1 score of the n-gram dev was 0.54

3g. A computer screen shot of a program

Description automatically generatedA screenshot of a web page

Description automatically generated

The F1 score is a bit higher than I expected, given the previous L1 F1 score, I was expecting something around 0.73

Written Exercises below.

**Written Exercises Below**

A white sheet of paper with text

Description automatically generated

A paper with writing on it

Description automatically generated

