Step #4
649 updates and 11 epochs during training
Training error rate: 0.0
Validation error rate: 0.025

Step #5
Highest weight words:
['click', 'remov', 'pleas', 'sight', 'pai', 'we', 'free', 'death', 'your', 'market', 'car', 'name']
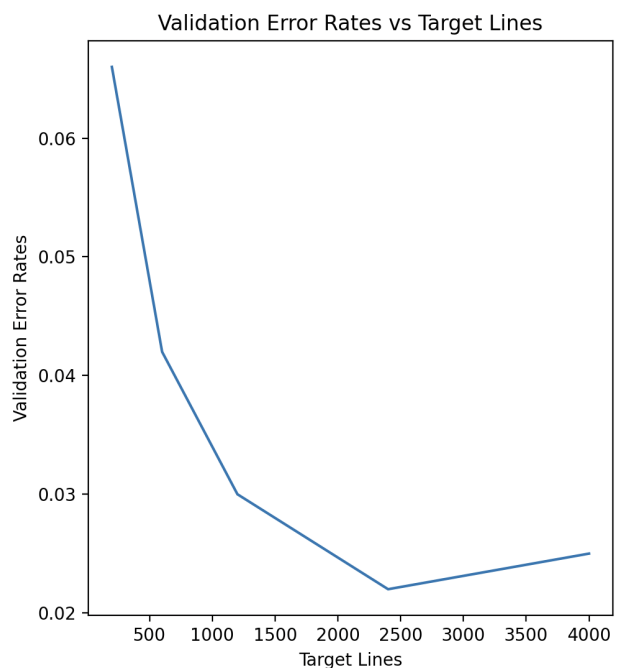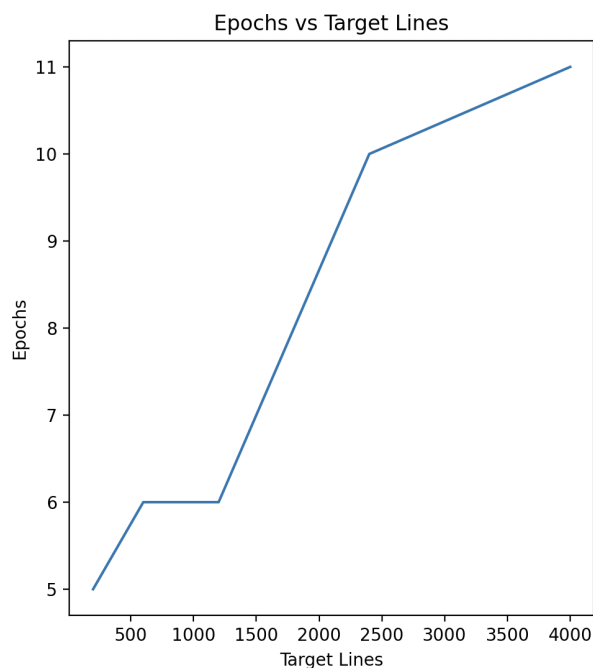Lowest weight words:
['wrote', 'but', 'which', 'view', 'would', 'reserv', 'there', 'seem', 'still', 'prefer', 'head', 'rob']

Step #6
target lines: [200, 600, 1200, 2400, 4000]
epochs: [5, 6, 6, 10, 11]
validation error rates: [0.066, 0.042, 0.03, 0.022, 0.025]

Step #7
Using Training data 4000 lines
max epochs: [10, 15, 20, 11]
validation error rates: [0.025, 0.025, 0.025, 0.025]
Limiting the epochs from 11 to 10 didn't change the validation error rate

Step #8
Best hyperparameters found! Validation error rate: 0.019
Best error rate (0.019)! thresh, max_epochs: 100 , 15

Step #9
Checking best thresh 100 and max_epoch 15 ...
Epochs: 15
Updates: 763
Error rate after testing the best w with the test data is 0.025 .

Step #10
In machine learning, it's crucial to have distinct training, validation, and test sets to ensure that the model generalizes well to new, unseen data.

The training set is the data on which the model is initially fitted. The training set helps the model learn the underlying patterns and relationships between the features and the target variable.
The validation set is used to fine-tune the model parameters and to provide an unbiased evaluation of the model during the training phase. Without a validation set, it's easy to create a model that performs well on the training data but poorly on unseen data (overfitting).
After training and validation, the test set provides the final, unbiased performance metric for the model. It's essential that the test set is entirely separate from the training and validation sets to give an accurate measure of how well the model will perform on never-before-seen data (assuming that it is from the same distribution).

By having these three distinct sets, we can better ensure that the model not only learns the most predictive features but also generalizes well to new data.