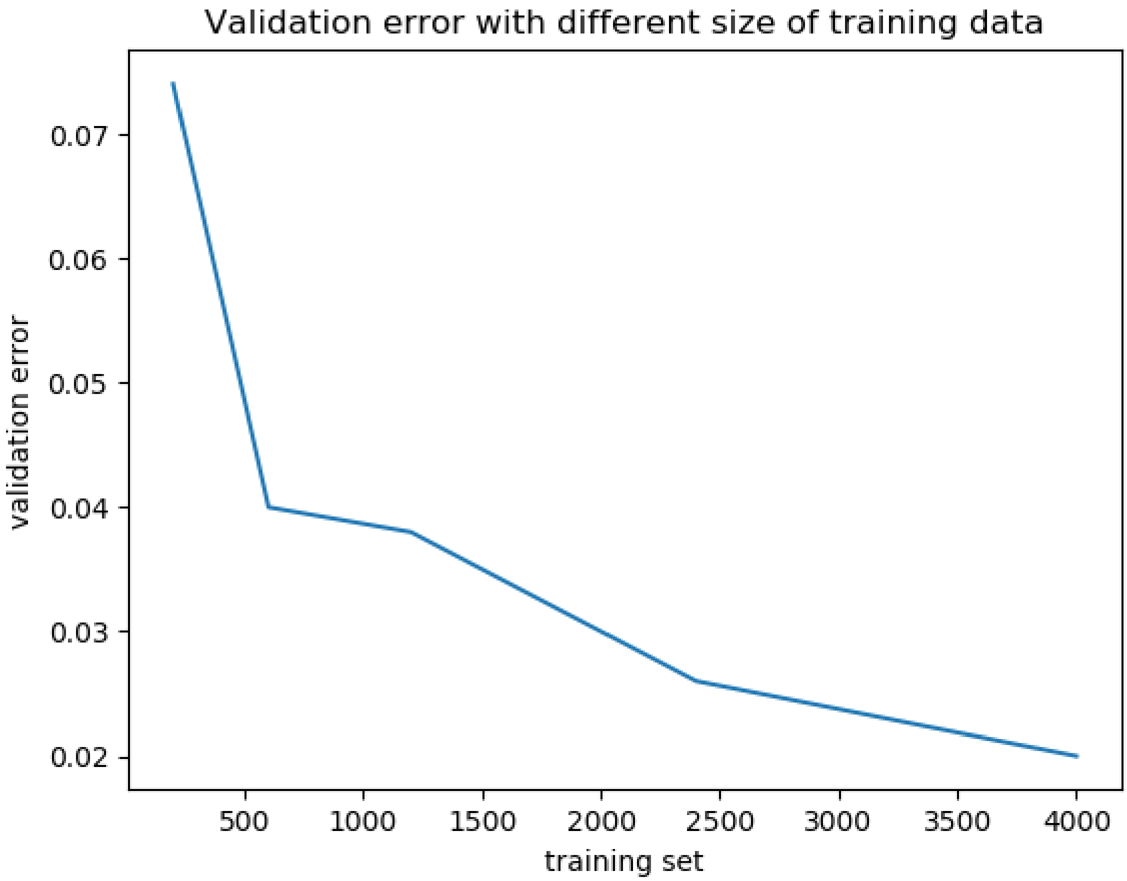
HW1

Program 1

- 1.3
  After setting bias, it made `430` mistakes in total.
  The error rate on training set is `0` . And it is `0.02(2%)` on validation set.
- 1.4
  Postive words:
  ```
  most positive weights are:
  remov
  best
  from
  take
  lost
  numbercnumb
  captur
  t
  have
  program
  some
  how
  ```
  Negative words:
  ```
  most negative weights are:
  what
  iso
  matthia
  few
  log
  discuss
  found
  transpar
  se
  file
  sponsor
  also
  ```
- 1.5


Validation error with different size of training data
validation error
0.07
0.06
0.05
0.04
0.03
0.02
500
1000
1500
2000
2500
3000
3500
4000
training set

- 1.6

- 1.7

  Set a max iteration to 30.
- 1.8

From the grpah, we can tell the best config is X = 27.
And the error rate on the whole test-set is 0.012(1.2%).

- 1.9
  38 features(37 words plus a bias). Error rate is 0.169(16.9%).
  It is not separable. With such a huge number, we can only get a vector that describes the most common words among all email types.
  Also, such a X also shows a huge validation error.
- 1.10
  By using the validatoin set, we can tune the hyper-parameter in our model like the `x` in our case.
  Also, bescuase the model has no prior-knowledge about the test dataset. We can get a more accurate benchmark on the test dataset.

Problem 2

Let's take two arbitrary points from the hyperplane x_1 amd x_2.

Substitute the two points to the hyperplane, we will get:

$$w \cdot x_1 + b = 0$$

$$w \cdot x_2 + b = 0$$

By doing subtrackion on above equations, we will get:

$$w \cdot (x_1 - x_2) = 0$$

which shows, vector w dot product any line in the plane equals zero and it means vector w is perpendicular to the hyperplane.

Problem 3

- Claim 2 :
  We have $w_t \Rightarrow w_{t-1} + x_{t-1}y_{t-1}$
  Therefor, $w_t \cdot w^* = (w_{t-1} + x_{t-1}y_{t-1})w^* = w_{t-1}w^* + y_{t-1}x_{t-1}w^*$.
  We know that $y_i w^* x_i \geq \gamma$, then we know $w_t w^* = w_{t-1}w^* + y_{t-1}x_{t-1}w^* \geq w_{t-1}w^* + \gamma$.
  Thus, we know $w_t w^* \geq \gamma + w_{t-1}w^*$;
  Similarly, we know $w_{t-1}w^* \geq \gamma + w_{t-2}w^*$
  By substituting them iteratively, we will have $w_t \cdot w^* \geq M_t\gamma$
  Claim 2 proof done.
- Claim 3:
  Given Claim2 & Claim3, we have:

$$\|w_t\|^2 \leq M_t R^2$$

$$w_t \cdot w^* \geq M_t\gamma$$

For second claim:

$$w_t \cdot w^* \geq M_t\gamma \Rightarrow \|w_t\|\|w^*\|cos\theta \geq M_t\gamma$$

Since we know by defination, $\|W^*\| = 1$,

$$w_t \cdot w^* \geq M_t\gamma \Rightarrow \|w_t\|cos\theta \geq M_t\gamma \Rightarrow \|w_t\|^2 cos^2\theta \geq {M_t}^2\gamma^2$$

We know that $\|w_t\|^2 \geq \|w_t\|^2 cos^2\theta$,

$$M_t R^2 \geq {M_t}^2 \gamma^2 \Rightarrow M_t \leq \frac{R^2}{\gamma^2}$$

Problem 4

By calculation, we know the R^2 for $X = 26$ is 1114. Also, we know the $M_t = 430$.
Therefore, we get $\gamma^2 \leq \frac{R^2}{M_t} \Rightarrow 0 \leq \gamma \leq 1.61$