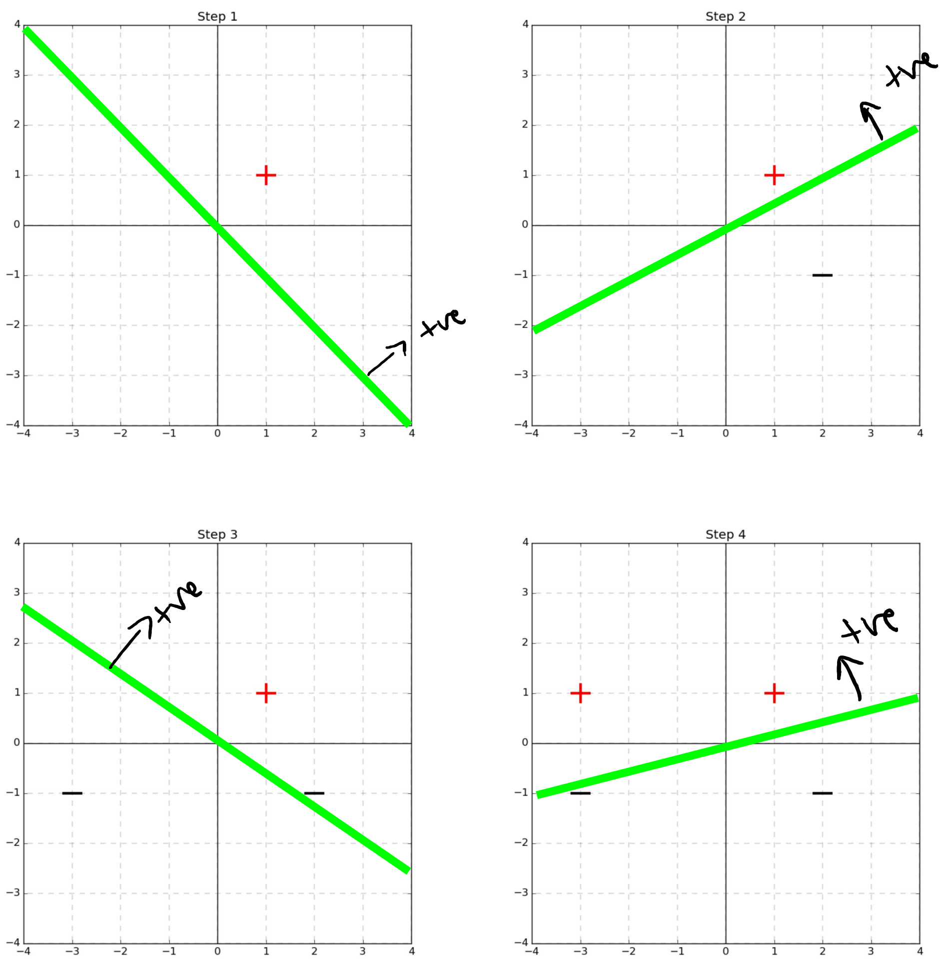
1. (10 points) Starting with $\mathbf{w} = [0 \quad 0]^T$, use the perceptron algorithm to learn on the data points in the order from top to bottom. Show the perceptron's linear decision boundary after observing each data point in the graphs below. Be sure to show which side is classified as positive.

$w^{(1)} = (1,1)$

$w^{(2)} = (-1,2)$

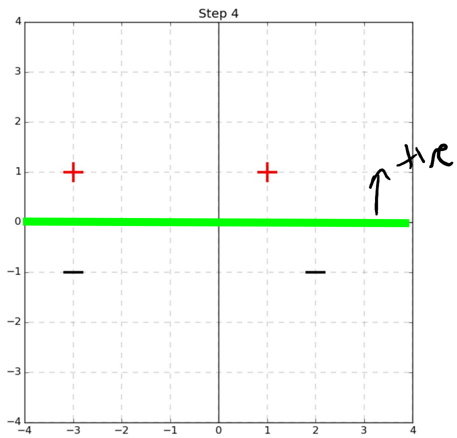


$w^{(3)} = (2,3)$

$w^{(4)} = (-1,4)$

2. (5 points) Does our learned perceptron maximize the margin between the training data and the decision boundary? If not, draw the maximum-margin decision boundary on the graph below.

No.
The maximum boundry would be the following



3. (7 points) Assume that we continue to collect data and train the perceptron. If all data we see (including the points we just trained on) are linearly separable separable with margin $\gamma = 0.5$ and have maximum norm $||\mathbf{x}_t|| \leq 5$, what is the maximum number of mistakes we can make on future data?

maximum mistakes $= \left(\frac{R}{\gamma}\right)^2$

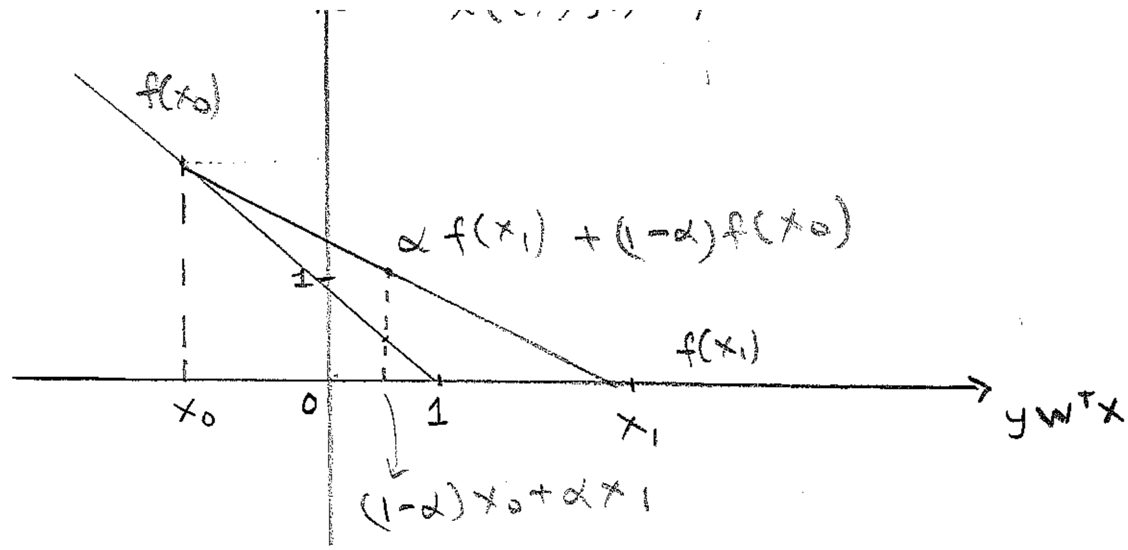
maximum norm $||x_t|| \leq 5$

$\Rightarrow R = 5$

and we know $\gamma = 0.5$

maximum mistakes $= \left(\frac{5}{0.5}\right)^2 = 100$

100



A function is convex is for all x_0, x_1 and $0<\alpha<1$

$$f((1-\alpha)x_0 + \alpha x_1) \leqslant (1-\alpha)f(x_0) + \alpha f(x_1)$$

As we can see from the image above the condition is followed everywhere.

→ if $x_0 < 1$ and $x_1 > 1$ we can see LHS is $<$ RH

→ if $x_0 < 1$ and $x_1 < 1$ or $x_0 > 1$ and $x_1 > 1$ LHS $=$ RHS

Therefore, we prove that the loss$((x, y), w)$ is convex.

CSE 446: Homework 3

ayush29f@cs.washington.edu
Ayush Saraf

Problem 2.2

We are interested in the hinge losses for the correctly classified examples, which is given by the following loss function:-

$$l((x,y),w) = max[0, 1 - yw^T x]$$

For all correctly classified examples x_i where $\bar{x_i}$ is the projection on the boundry,

$$y_i w^T x_i \in [y_i w^T \bar{x_i}, \infty]$$

$$y_i w^T x_i \in [0, \infty]$$

$$-y_i w^T x_i \in [-\infty.0]$$

$$1 - y_i w^T x_i \in [-\infty.1]$$

$$max(0, 1 - y_i w^T x_i) \in [0.1]$$

Therefore,

$$loss((x_i, y_i), w) \in [0, 1]$$

Problem 2.3

Let $M(w)$ be the number of mistakes made by our model in our dataset of N items. So the total loss could be written as,

$$\sum_{i=1}^{N} max[0, 1 - yw^T x_i] = A + B$$

Let the total loss made by mistakes be A and total loss made by correct predictions be B

$$A = \sum_{i \in mistakes} max[0, 1 - yw^T x_i]$$

$$B = \sum_{i \notin mistakes} max[0, 1 - yw^T x_i]$$

Since we know that

$$\forall_{i \in mistakes} max[0, 1 - yw_T x_i] \geq 1$$

$$A \in [M(w), \infty)$$

And we know,

$$\forall_{i \notin mistakes} max[0, 1 - yw_T x_i] \in [0, 1)$$

$$B \in [0, N - M(w))$$

Therefore,

$$A + B \in [M(w), \infty)$$

$$\frac{A+B}{N} \in [\frac{1}{N}M(w), \infty)$$

Therefore,

$$\frac{1}{N}M(w) \leq \frac{1}{N}\sum_{i=1}^{N} max[0, 1 - yw^T x_i]$$

Problem 3.1

A linearly separable boundary equation looks like the following,

$$\mathbf{w^T}\phi(\mathbf{x}) + w_0 = 0$$

Too show that a circular boundary with a center (a, b) and radius r is linearly separable we have to show that the equation of the circle could follow the above equation.

$$(x[1] - a)^2 + (x[2] - b)^2 - r^2 = 0$$

$$x[1]^2 - 2ax[1] + a^2 + x[2]^2 - 2bx[2] + b^2 - r^2 = 0$$

$$x[1]^2 + x[2]^2 - 2bx[2] - 2ax[1] + a^2 + b^2 - r^2 = 0$$

This could be rewritten as,

$$\phi(\mathbf{x}) = (x[1], x[2], x[1]^2, x[2]^2)$$

$$\mathbf{w} = (-2a, -2b, 1, 1)$$

$$w_0 = a^2 + b^2 - r^2$$

Therefore, we showed that the circular equation is linearly separable in the space $\phi(\mathbf{x})$

Problem 3.2

A linearly separable boundary equation looks like the following,

$$\mathbf{w^T}\phi(\mathbf{x}) + w_0 = 0$$

Too show that a elliptical boundary is linearly separable we have to show that the equation of the circle could follow the above equation.

$$c(x[1] - a)^2 + d(x[2] - b)^2 - r^2 = 0$$

$$cx[1]^2 - 2acx[1] + a^2c + dx[2]^2 - 2bdx[2] + b^2d - r^2 = 0$$

$$cx[1]^2 + dx[2]^2 - 2acx[1] - 2bdx[2] + a^2c + b^2d - r^2 = 0$$

This could be rewritten as,

$$\phi(\mathbf{x}) = (x[1], x[2], x[1]^2, x[2]^2, x[1]x[2])$$

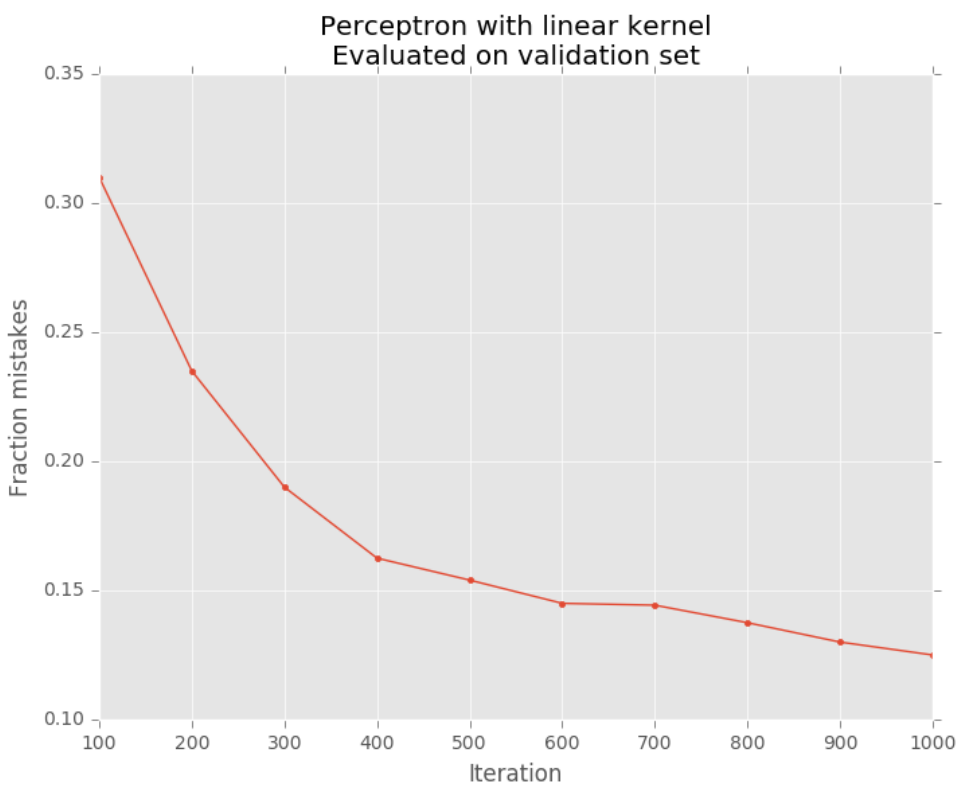
$$\mathbf{w} = (-2ac, -2bd, c, d, 0)$$

$$w_0 = a^2c + b^2d - r^2$$

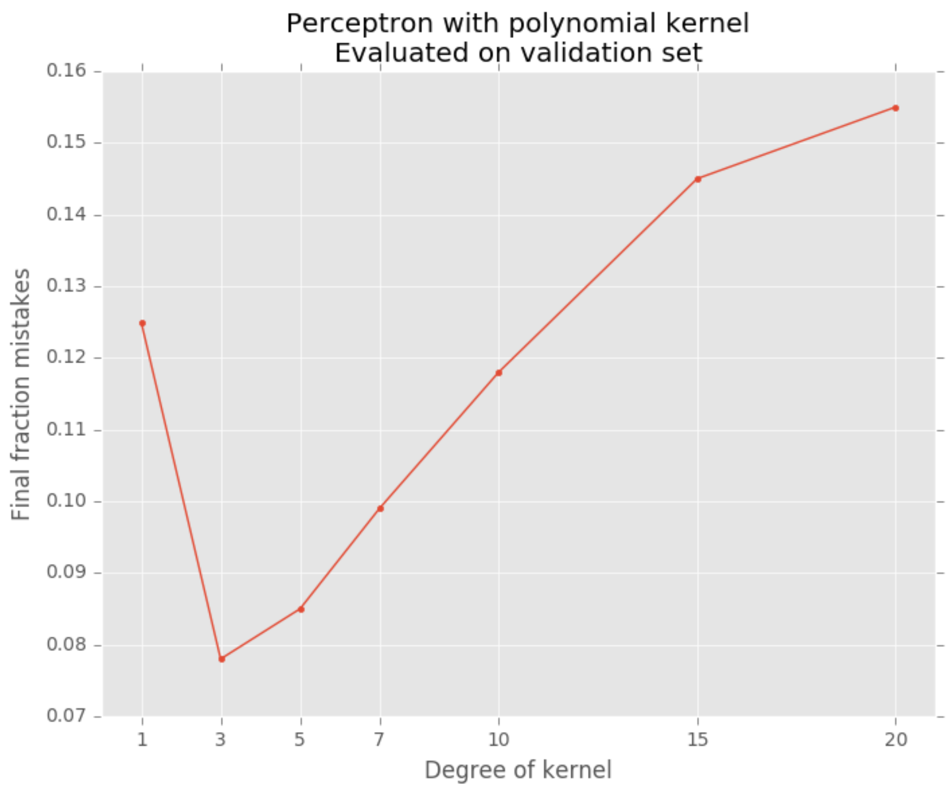
Therefore, we showed that the elliptical equation is linearly separable in the space $\phi(\mathbf{x})$

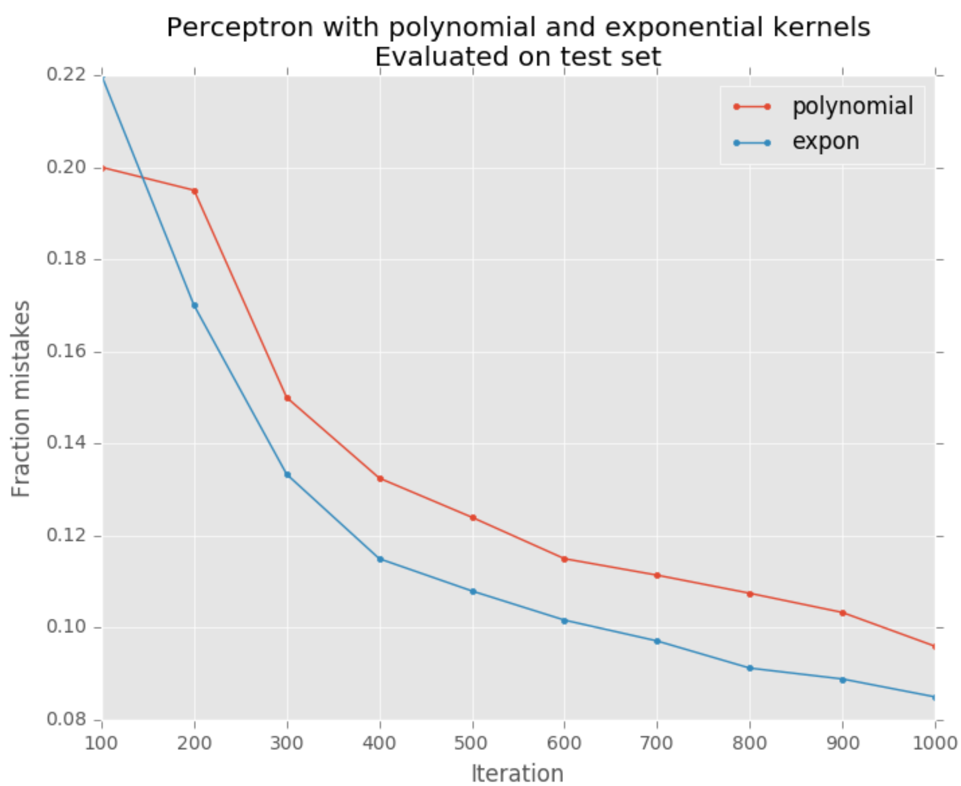
Problem 4.4

Looking at the first plot we can see that the Perceptron algorithm is getting better as we train the model with multiple steps since the fraction of mistakes are decreasing. Further, we use a polynomial kernel and check the validation accuracy with different dimension polynomial, it is in fact interesting to see that the validation accuracy stars to decrease after degree=3. We note that this is the most optimal degree and then train both polynomial with degree=3 and exponential kernel Perceptron and evaluate the accuracy on the test set. We can see as the that the exponential kernel easily outperforms the polynomial. This is in consistency with our expectations because an exponential kernel space is much larger than the polynomial one. This allows us to work in a much higher dimensional space and we find that the data is linearly separable in the higher dimensional space i.e. exponential



kernel.


Perceptron with polynomial kernel
Evaluated on validation set
Final fraction mistakes
Degree of kernel
0.16
0.15
0.14
0.13
0.12
0.11
0.10
0.09
0.08
0.07
1
3
5
7
10
15
20


Perceptron with polynomial and exponential kernels
Evaluated on test set
polynomial
expon
Fraction mistakes
Iteration
0.22
0.20
0.18
0.16
0.14
0.12
0.10
0.08
100
200
300
400
500
600
700
800
900
1000