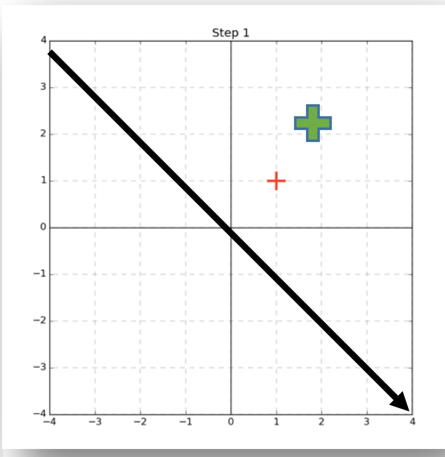
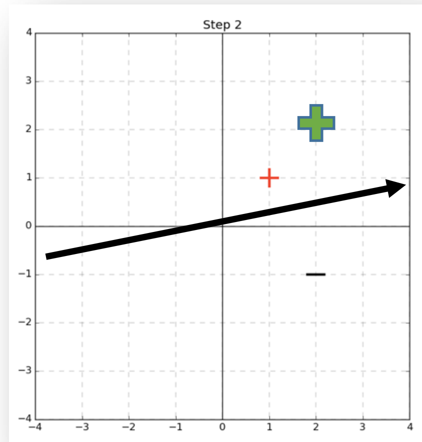
Michael Lee
HW #3

Perceptron   Note:  Means that the side of the graph was considered positive

1.





1st Pass:

$sign\left(dot(w^T, [1,1])\right) = sign(0) = 0 \;!\!= 1$

Algorithm was wrong so **w** is updated:

$\boldsymbol{w}^1 = \boldsymbol{w}^0 + y_0 * \boldsymbol{x}_0$

$\boldsymbol{w}^1 = \begin{bmatrix}0\\0\end{bmatrix} + 1 * \begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}1\\1\end{bmatrix}$

The boundary is therefore: x[2] = -x[1]

2nd Pass:

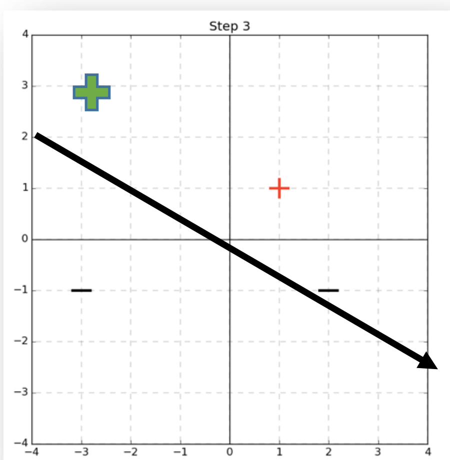
$sign\left(dot(w^T, [2-1]])\right) = sign(1) = 1 \,!= -1$

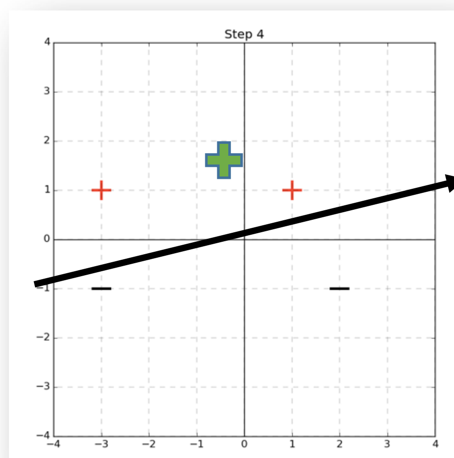
Algorithm was wrong so **w** is updated:

$\boldsymbol{w}^2 = \boldsymbol{w}^1 + y_1 * \boldsymbol{x}_1$

$\boldsymbol{w}^1 = \begin{bmatrix}1\\1\end{bmatrix} + -1 * \begin{bmatrix}2\\-1\end{bmatrix} = \begin{bmatrix}-1\\2\end{bmatrix}$

The boundary is therefore: x[2] = 0.5 * x[1]





3rd Pass:

$sign\left(dot(w^T, [-3,1])\right) = sign(6) = 1$!= -1

Algorithm was wrong so **w** is updated:

$\boldsymbol{w}^3 = \boldsymbol{w}^2 + y_2 * \boldsymbol{x}_2$

$\boldsymbol{w}^3 = \begin{bmatrix}-1\\2\end{bmatrix} + -1 * \begin{bmatrix}-3\\-1\end{bmatrix} = \begin{bmatrix}2\\3\end{bmatrix}$

The boundary is therefore: x[2] = -2/3x[1]

Final Pass:

$sign\left(dot(w^T, [-3,1])\right) = sign(-3) = -1 \;!\!= 1$

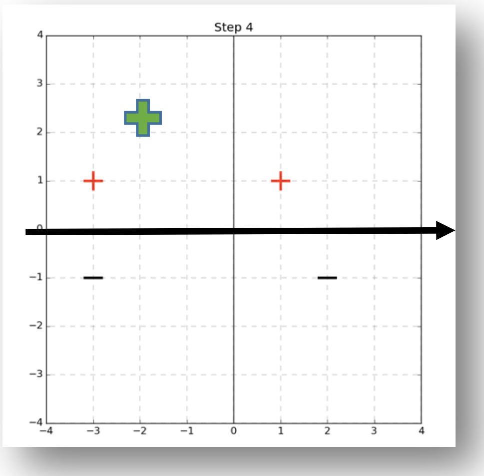
Algorithm was wrong so **w** is updated:

$\boldsymbol{w}^4 = \boldsymbol{w}^3 + y_3 * \boldsymbol{x}_3$

$\boldsymbol{w}^4 = \begin{bmatrix}0\\0\end{bmatrix} + 1 * \begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}-1\\4\end{bmatrix}$

The boundary is therefore: x[2] = 0.25x[1]

2. Our learned perceptron does not draw the maximized margin. We see that if we pick the line x[2]=0 we get a maximized margin of 1.

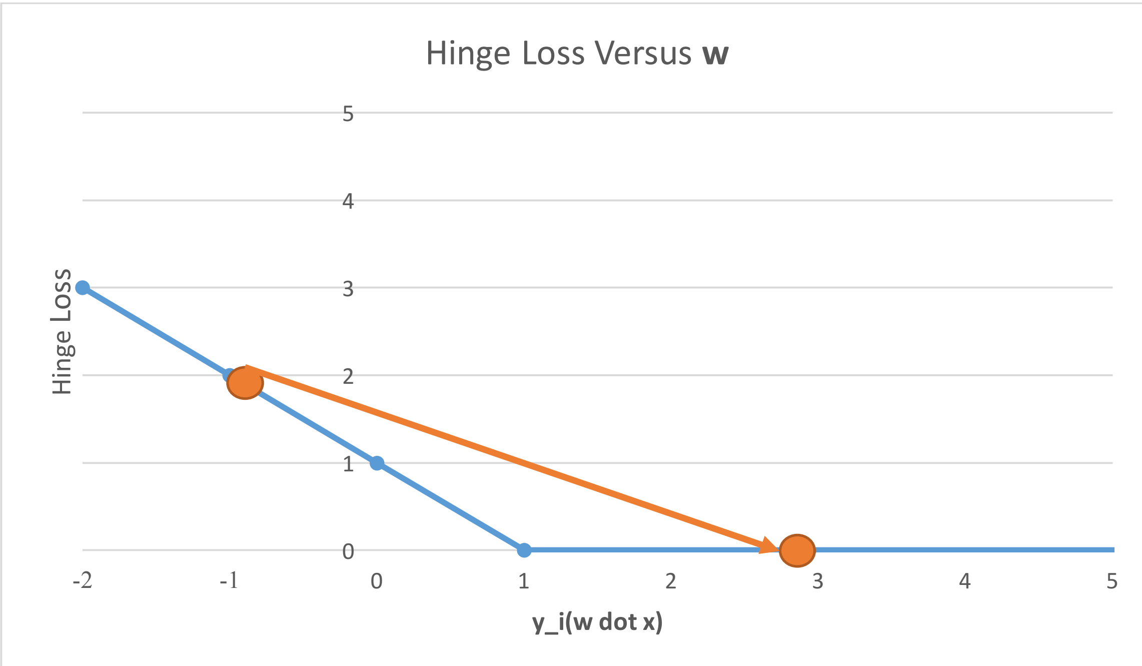


3.

Max Mistakes = $\left(\frac{\gamma}{R}\right)^2 = \left(\frac{5}{0.5}\right)^2 = 100$

SVMs: Hinge Loss and Mistake Bounds

1.



Here we see that the function is convex because if we take any two points (as shown in the orange) that any line we draw will be greater than or equal to the minimum of 0.

2. For a fixed $\boldsymbol{w}$ that gives the correct solution we take on the values [0,1]. This is because we know the term $y\boldsymbol{w}^T\boldsymbol{x}$ will necessarily be positive as a result of **w** giving the correct guess. Hence max[0, 1- $y\boldsymbol{w}^T\boldsymbol{x}$] will be equivalent to max[0 to 1 – α] where $\alpha = y\boldsymbol{w}^T\boldsymbol{x}$ and $\alpha > 0$. The maximum value would then occur at $\alpha = 0$ which would make the function return 1. The minimum would occur at 0. So the range is [0,1]

3. We consider the bound in two cases:

-In the best case M(w) equals 0 and the classification is therefore 0.

So we have:

$$0 \leq \frac{1}{N}\sum_{i=1}^{N} \max\,[0,1 - y\boldsymbol{w}^T\boldsymbol{x}]$$

We can observe that the minimum value of $\frac{1}{N}\sum_{i=1}^{N} \max\,[0,1 - y\boldsymbol{w}^T\boldsymbol{x}]$ would be for the function $\max\,[0,1 - y\boldsymbol{w}^T\boldsymbol{x}]$ to return 0. And hence we would get $\sum_{i=1}^{N}, 0$ which evaluates to 0, so the resulting equation would be 0<=0 and the upper bound would hold.

[continued below]

-In the worst case M(w) equals N we have:

$$\frac{N}{N} \leq \frac{1}{N}\sum_{i=1}^{N} \max\,[0,1-y\boldsymbol{w}^T\boldsymbol{x}] \rightarrow 1 \leq \frac{1}{N}\sum_{i=1}^{N} \max\,[0,1-y\boldsymbol{w}^T\boldsymbol{x}]$$

Note: In the prior problem we showed that if $\boldsymbol{w}$ is a correct solution then $\max[0,1-y\boldsymbol{w}^T\boldsymbol{x}] \in [0,1]$. If $\boldsymbol{w}$ is an incorrect solution, we can say that this function is necessarily greater than or equal to one. This is because if **w** is incorrect we know $y\boldsymbol{w}^T\boldsymbol{x}$ will be a negative number and hence the function $\max[0,1-y\boldsymbol{w}^T\boldsymbol{x}]$ will be equivalent to $\max[0,1-\beta]$ where $\beta = y\boldsymbol{w}^T\boldsymbol{x}$ and $\beta \leq 0$. Therefore, the function will return the number a number greater than or equal to one.

So:

$1 \leq \frac{1}{N}\sum_{i=1}^{N} \kappa$ --$\kappa = \max[0,1-y\boldsymbol{w}^T\boldsymbol{x}]$ *and as proved above* $\kappa \geq 1$

$1 \leq \dfrac{1}{N} N * \kappa$

$1 \leq \kappa$

$1 \leq 1$

Since we have considered both cases of the inequality we can assert that it holds true that $\frac{M(\boldsymbol{w})}{N} \leq \frac{1}{N}\sum_{i=1}^{N} \max\,[0,1-y\boldsymbol{w}^T\boldsymbol{x}]$.

<u>Kernel Functions and Linear Separability</u>

1. $(\boldsymbol{x}[1]-a)^2 + (\boldsymbol{x}[2]-b)^2 - r^2 = 0$

   *We multiply out terms to get:
   $\boldsymbol{x}[1]^2 - 2\boldsymbol{x}[1]a + a^2 + \boldsymbol{x}[2]^2 - 2\boldsymbol{x}[2]b + b^2 - r^2 = 0$

We recognize that this equation represents a four dimensional plane, rewriting the equation we see:

$$[-2a, -2b, 1,1]\begin{bmatrix} x[1] \\ x[2] \\ x[1]^2 \\ x[2]^2 \end{bmatrix} + b^2 + a^2 - r^2 = 0$$

Since the vector has linear coefficients we can assert that the equation is linear. Therefore, in four dimensions we see there is always a linear decision boundary.

2. $c * (\boldsymbol{x}[1]-a)^2 + d * (\boldsymbol{x}[2]-b)^2 - 1 = 0$

   *We multiply out terms to get:
   $c\boldsymbol{x}[1]^2 - 2\boldsymbol{x}[1]ac + ca^2 + d\boldsymbol{x}[2]^2 - 2\boldsymbol{x}[2]bd + db^2 - 1 = 0$

This can be rewritten as:

$$[-2ac, -2bd, c, d, 0]\begin{bmatrix} x[1] \\ x[2] \\ x[1]^2 \\ x[2]^2 \\ x[1]x[2] \end{bmatrix} + b^2 + a^2 - r^2 = 0$$

We recognize that this equation represents a four dimensional plane because the vector represented only has linear coefficients, that is: $[-2ac, -2bd, c, d, 0]$. Therefore, in a five dimensional feature space there is always a linear decision boundary.

<u>Programming Question- Wrap up</u>

The 3rd degree polynomial performed the best and the exponential kernel outperformed the polynomial kernel. The results matched my expectations, given the wider range of severability that an exponential will spread the data with it seemed like it would form a better decision boundary.