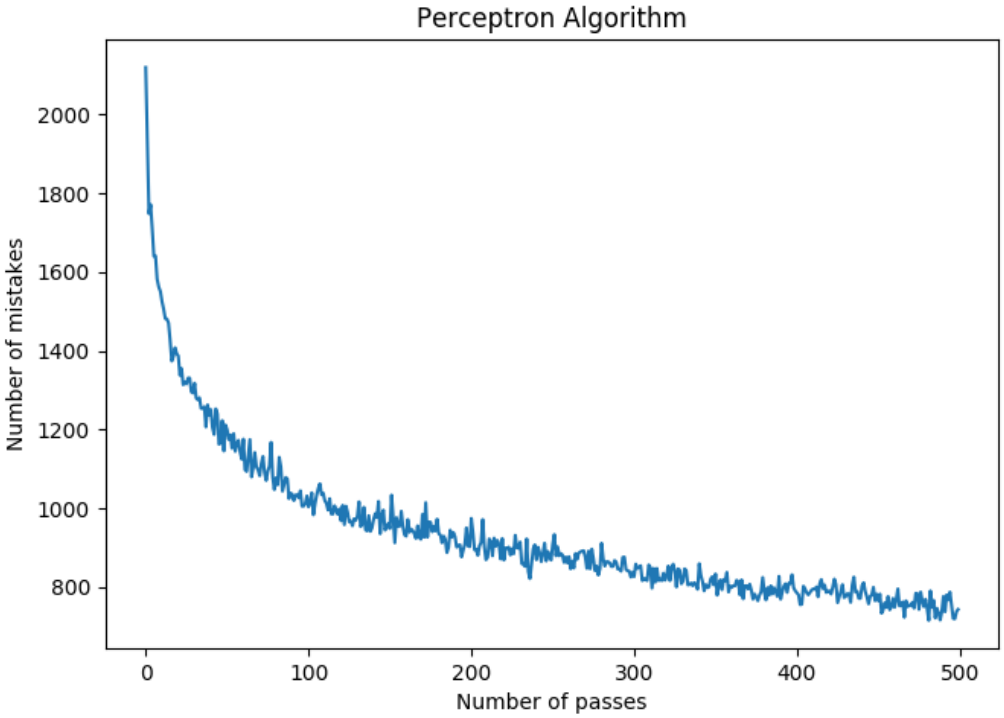
Yerbol Aussat

Perceptron Leaning



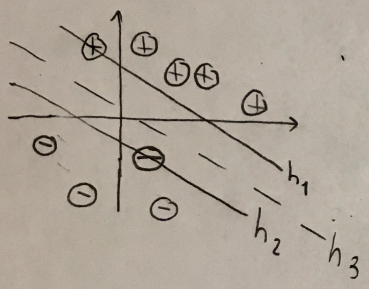
$$\begin{cases} \langle \underset{\sim}{x}_i, \underset{\sim}{w}^* \rangle + b^* \geq 0 & \text{for } y_i = 1 \\ \langle \underset{\sim}{x}_i, \underset{\sim}{w}^* \rangle + b^* < 0 & \text{for } y_i = -1 \end{cases}$$

where $\langle \underset{\sim}{x}_i, \underset{\sim}{w}^* \rangle + b^*$ is a hyperplane h_1.

If we move hyperplane h_1 in the direction of its norm, we can get another hyperplane h_2 s.t.

$$\begin{cases} \langle \underset{\sim}{x}_i, \underset{\sim}{w}^* \rangle + (b^* + \delta) > 0 & \text{for } y_i = 1 \\ \langle \underset{\sim}{x}_i, \underset{\sim}{w}^* \rangle + (b^* + \delta) < 0 & \text{for } y_i = -1 \end{cases}$$

, where $\delta > 0$ is an "offset" just enough to "touch" the first point in negative halfspace



. Therefore, there must exist the hyperplane with the offset of $\frac{\delta}{2}$, and for this plane the following conditions hold:

$$\begin{cases} \langle \underset{\sim}{x}_i, \underset{\sim}{w}^* \rangle + (b^* + \frac{\delta}{2}) > 0 & \text{for } y_i = 1 \\ \langle \underset{\sim}{x}_i, \underset{\sim}{w}^* \rangle + (b^* + \frac{\delta}{2}) < 0 & \text{for } y_i = -1 \end{cases}$$

Thus, by construction, we demonstrated that there exist $\underset{\sim}{w}^{**} = \underset{\sim}{w}^*$ and $b^{**} = b^* + \frac{\delta}{2}$ such that

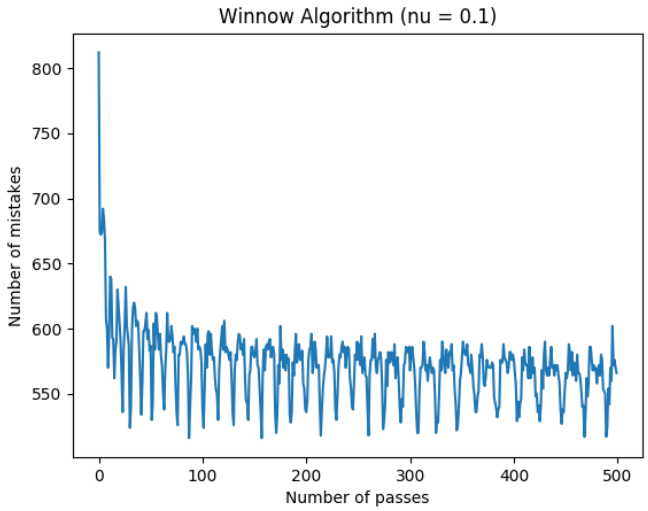
$$\begin{cases} \langle \underset{\sim}{x}_i, \underset{\sim}{w}^{**} \rangle + b^{**} > 0 \\ \langle \underset{\sim}{x}_i, \underset{\sim}{w}^{**} \rangle + b^{**} < 0 \end{cases}$$

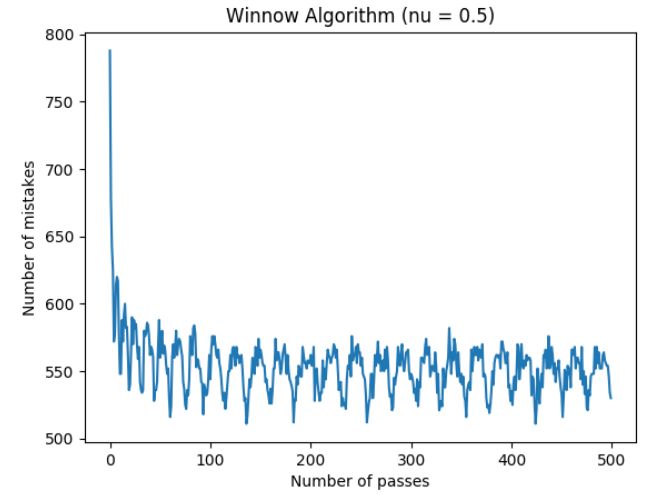
Yerbol Aussat

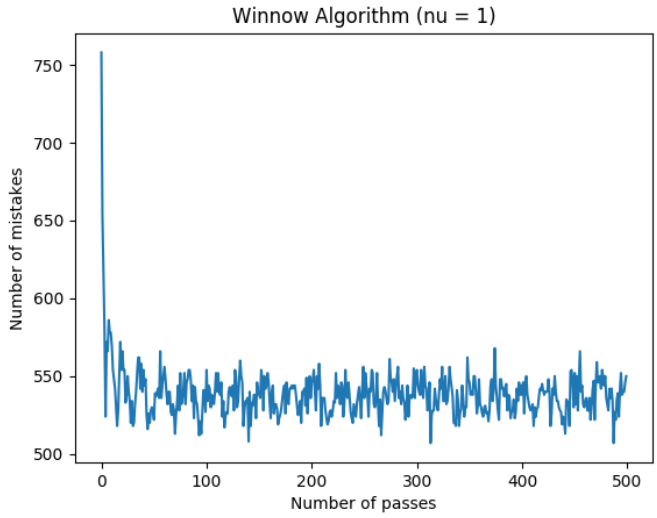
Winnow Learning

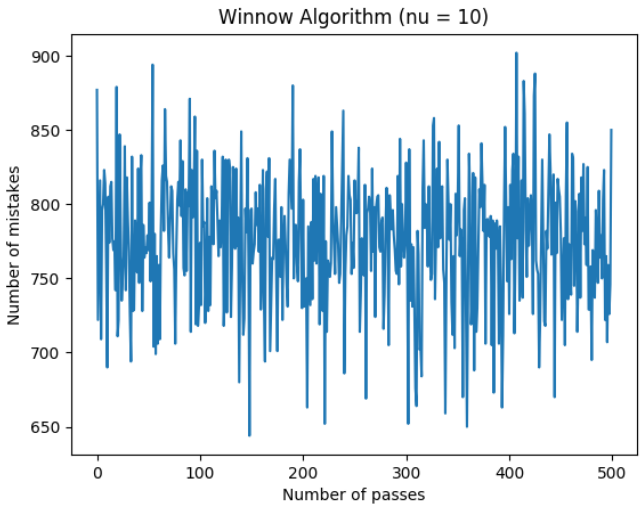
We can define two weight vectors as $w^+ = \max(w, 0)$ and $w^- = \max(-w, 0)$. According to this definition both weight vectors are greater or equal to zero. It can be seen that $w = w^+ - w^-$ Therefore, $\langle w, x \rangle = \langle w^+ - w^-, x \rangle = \langle w^+, x \rangle - \langle w^-, x \rangle$. Thus the non-negativity assumption is not violated, and $(w^+ - w^-)$ can take positive and negative values.

X normalized:
Below are number of mistakes vs number of passes for winnow algorithm for different values of η.







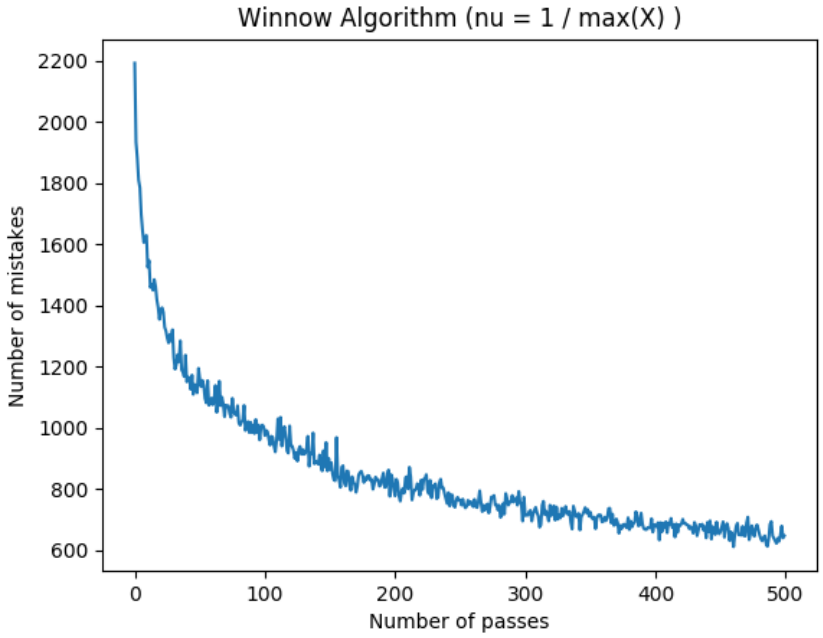


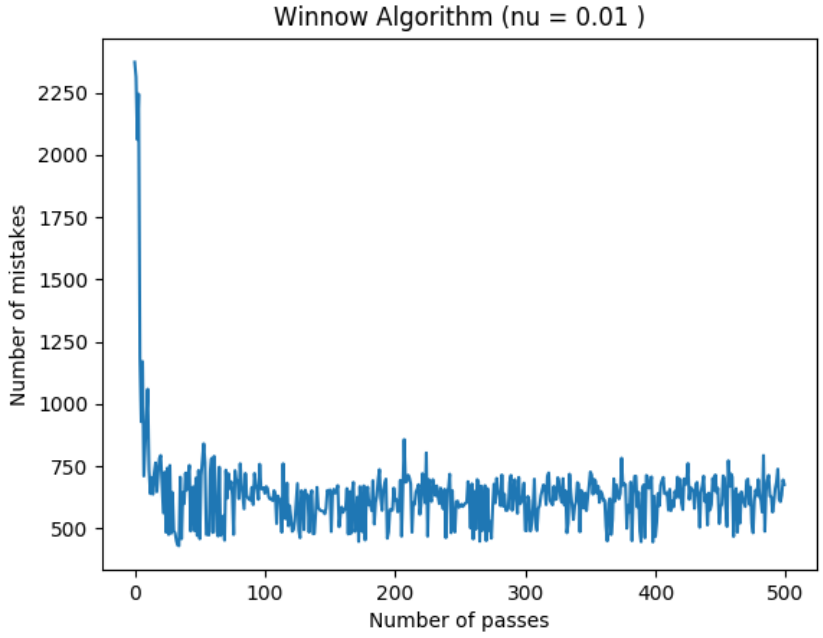
As we can see, the algorithm demonstrates the best performance when the step size $\eta = 1$, which is exactly one over the maximum absolute value of X. When η is too

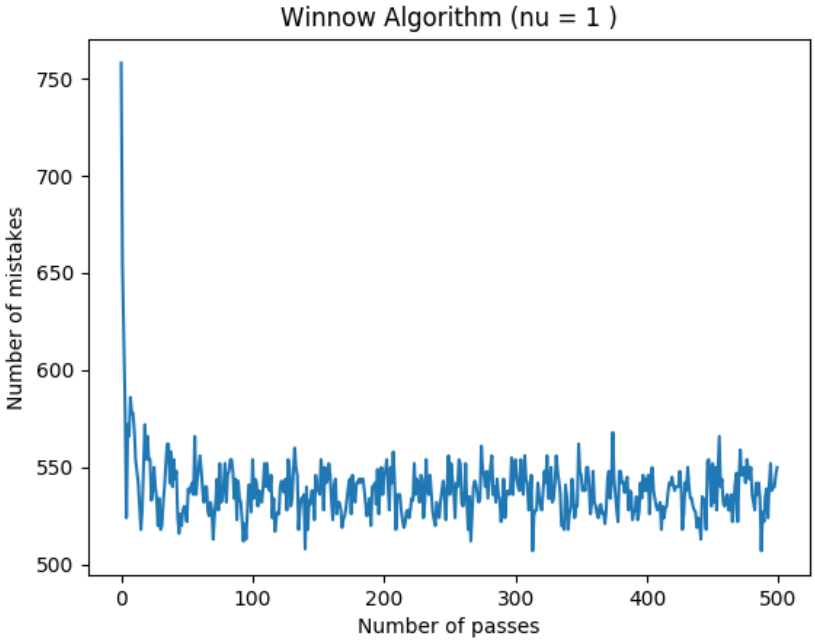
high, the algorithm doesn't converge, and when η is too low the fluctuations around are higher.

X not normalized:
Interestingly, much smoother results were obtained without normalization.

Thus, when η equals one over the maximum element of X, which is `6.312e-05`, algorithm has the following performance:

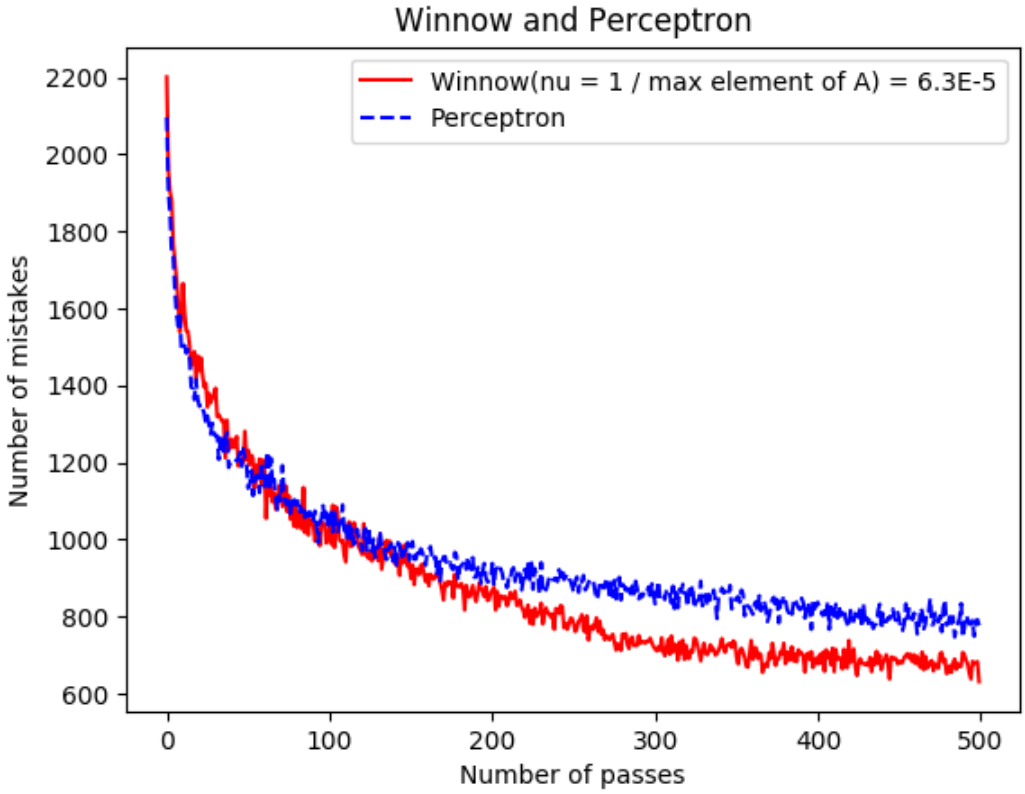





Winnow Algorithm (nu = 1)
Number of mistakes
750
700
650
600
550
500
0
100
200
300
400
500
Number of passes

Yerbol Aussat

Perceptron vs. Winnow



Based on the experiments, it was observed that Winnow converges faster than perceptron. Perceptron doesn't have a step size that needs to be optimized, whereas for Winnow the step size needs to be adjusted. However, the value of one over the value of the maximum element of X seems to work well. Balanced version of Winnow algorithm was able to classify spam, whereas the original version didn't work. Overall, both algorithms are good for binary classification, and each of them has its advantages and disadvantages.