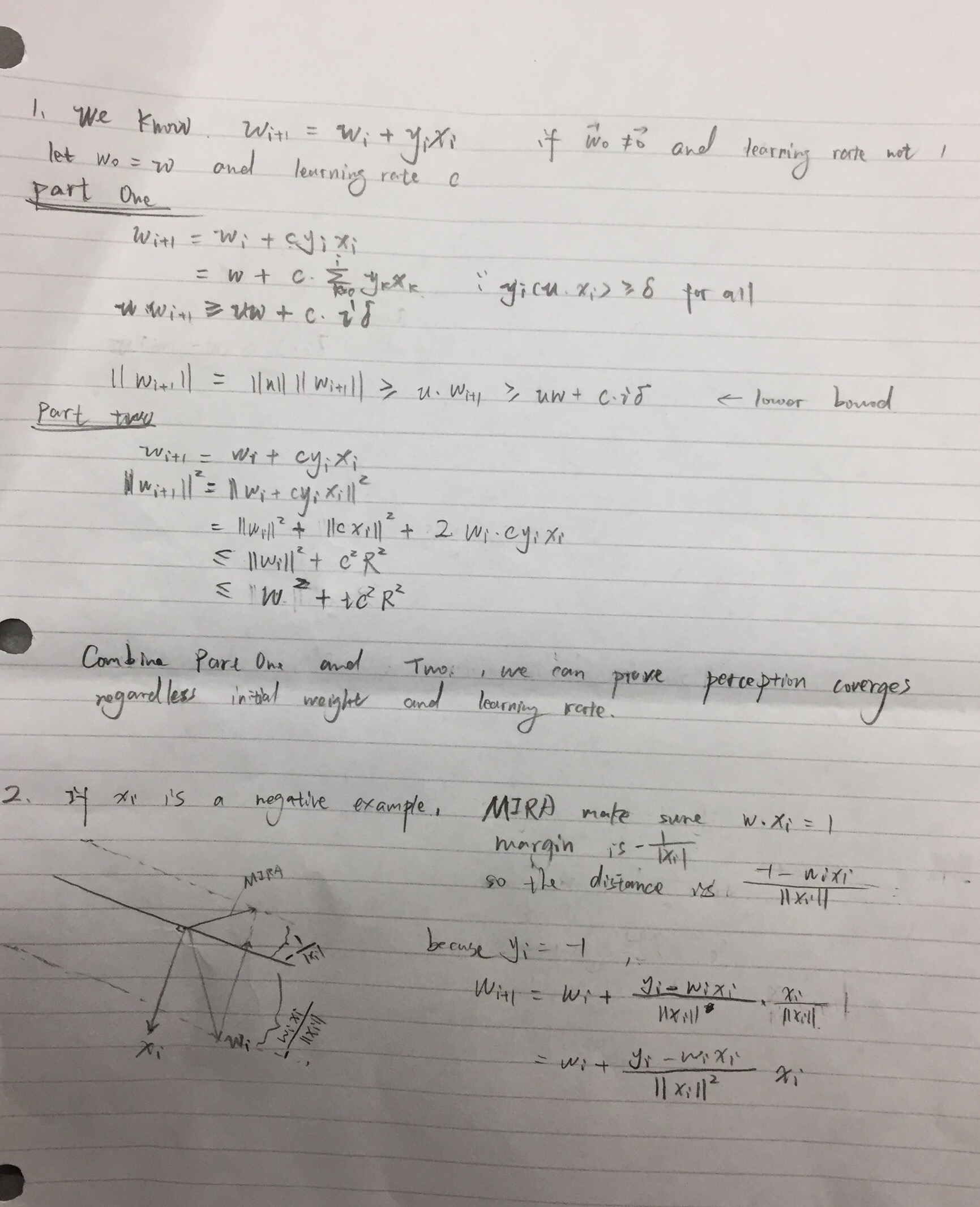
**CS534 Xiao Tan Exercise 1**



4.

After doing zero-mean and unit-variance, it will make data gradient descent converges much faster. Will make the graph easier to read.

5.

According to the naive implementations, it will do w’<-w’ + w every time, so that the running time will be DT times. However, if we do smart average perceptron, we only need to run the branch when it need to be updated. So, it will be run U times. U<<TD, so that smart will be less than naive.

DEBRIEF SECTION (required)

1. Did you work alone, or did you discuss with other students?

If the latter please write down their names.

Note: in general, only high-level discussions are allowed.

Discuss with Kaiwen Zheng, Haoyu Zhang

2. How many hours did you spend on this assignment?

8 hours

3. Would you rate it as easy, moderate, or difficult?

Difficult

4. Are the lectures too fast, too slow, or just in the right pace?

Too fast

5. Any other comments?

Notations in PowerPoint are not very clear. If it can be detailed will be better.