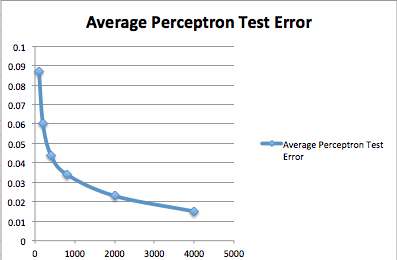
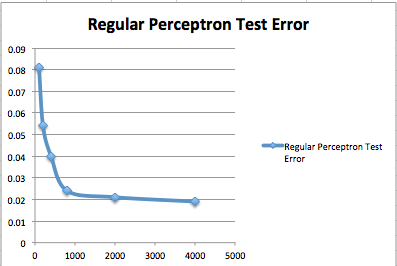
Jacob Preston

Machine Learning Homework #1

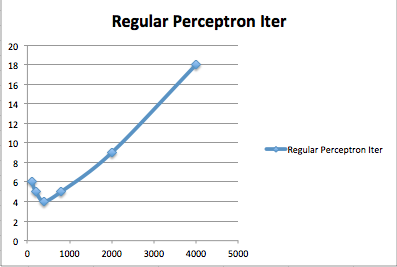
9/17/13

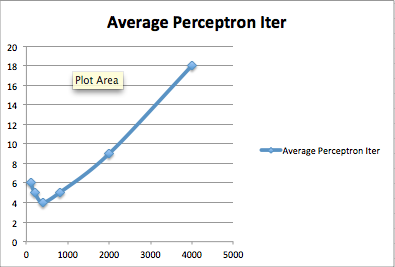
Write Up

1. It is important to create the validation set to be able to actually measure the results of your learning. Otherwise, if you really messed up, there would be no way of knowing and when you deploy your learning algorithm on unmarked data, it may predict completely wrong results
2. (In code)
3. (In code)
4. 429 mistakes are made in 19 iterations; Validation error is .019 or 1.9%
   1. Most positive:
      1. sight:20.0
      2. pleas:15.0
      3. nbsp:14.0
      4. our:14.0
      5. click:14.0
      6. remov:14.0
      7. guarante:14.0
      8. internet:13.0
      9. market:13.0
      10. most:12.0
      11. yourself:12.0
      12. present:12.0
      13. major:12.0
      14. deathtospamdeathtospamdeathtospam:12.0
      15. further:12.0
   2. Most negative
      1. wrote:-15.0
      2. prefer:-14.0
      3. i:-14.0
      4. part:-12.0
      5. version:-12.0
      6. re:-11.0
      7. on:-11.0
      8. run:-11.0
      9. review:-10.0
      10. url:-10.0
      11. p:-10.0
      12. recipi:-10.0
      13. comput:-10.0
      14. reserv:-10.0
      15. rpm:-10.0
5. (in code)



8.





9. (in code)

10.

1. 4000 average, 18 iter, learning rate = 1: .015
2. 4000 regular, 18 iter, learning rate = 1: .019
3. 4000 average, 15 iter, learning rate = 1: .016
4. 4000 regular, 15 iter, learning rate = 1:.018
5. 4000 average, 9 iter, learning rate = .25: .021
6. 4000 regular, 18 iter, learning rate = .25: .021

Final test scenario is using average perceptron and a max of 15 iterations and it gets a final test error of .0014