CS534 Machine Learning

Implementation #2
Fall 2019

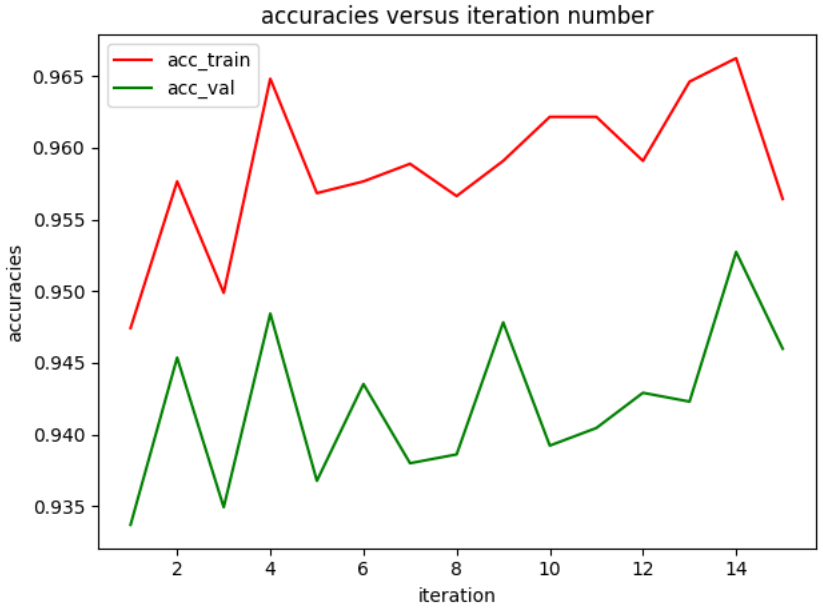
Oregon State of University

Instructor:
Fern, Xiaoli
Team Member:
Zhengqiang Yue (933-284-058) yuez@oregonstate.edu (34%)
Yufei Bai (933-315-539) baiyuf@oregonstate.edu (33%)
YuWen Tseng (933-652-910) tsengyuw@oregonstate.edu (33%)

PART 1: Online Perceptron

(a.)

Iteration	Accuracy for the train	Accuracy for the validation
1	0.947422	0.933702
2	0.957651	0.945365
3	0.949877	0.934929
4	0.964812	0.948435
5	0.956833	0.936771
6	0.957651	0.943524
7	0.958879	0.937999
8	0.956628	0.938613
9	0.959083	0.947821
10	0.962152	0.939227
11	0.962152	0.940454
12	0.959083	0.942910
13	0.964607	0.942296
14	0.966244	0.952732
15	0.956424	0.945979



The result shows that the train accuracy can't reach to 100%. Because the data is not linearly separable.

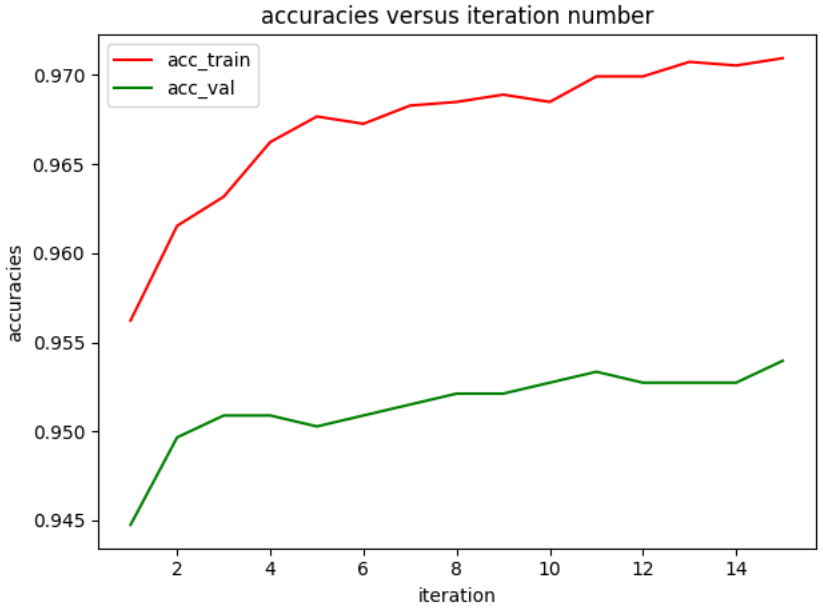
(b.)

From the above result, we get the best accuracy = 0.952732 when iteration = 14. The prediction set is in the "oplabel.csv".

Part2: Average Perceptron

(a.)

Iteration	Accuracy for the train	Accuracy for the validation
1	0.956219	0.944751
2	0.961538	0.949662
3	0.963175	0.950890
4	0.966244	0.950890
5	0.967676	0.950276
6	0.967267	0.950890
7	0.968290	0.951504
8	0.968494	0.952118
9	0.968903	0.952118
10	0.968494	0.952732
11	0.969926	0.953346
12	0.969926	0.952732
13	0.970745	0.952732
14	0.970540	0.952732
15	0.970949	0.953959



(b.)
Comparing to online perceptron, average perceptron leads to a curve with higher accuracy. The curve is also smoother because it uses average weight.

(c.)
Based on the result of validation accuracy, we choose iteration = 15. The prediction set is in the "aplabel.csv".

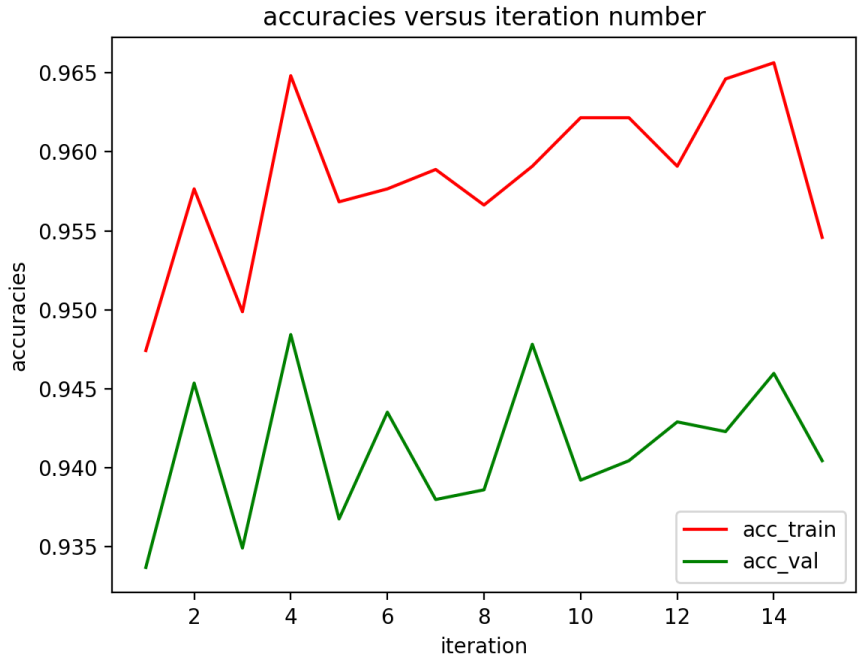
PART 3: Polynomial Kernel Perceptron

(a.1)

[P = 1]

Iteration	Accuracy for the train	Accuracy for the validation
1	0.947422	0.933702
2	0.957651	0.945365
3	0.949877	0.934929
4	0.964812	0.948435
5	0.956833	0.936771
6	0.957651	0.943524
7	0.958879	0.937999
8	0.956628	0.938613
9	0.959083	0.947821
10	0.947821	0.939227
11	0.962152	0.940454
12	0.959083	0.942910
13	0.964607	0.942296
14	0.965630	0.945979
15	0.954583	0.940454

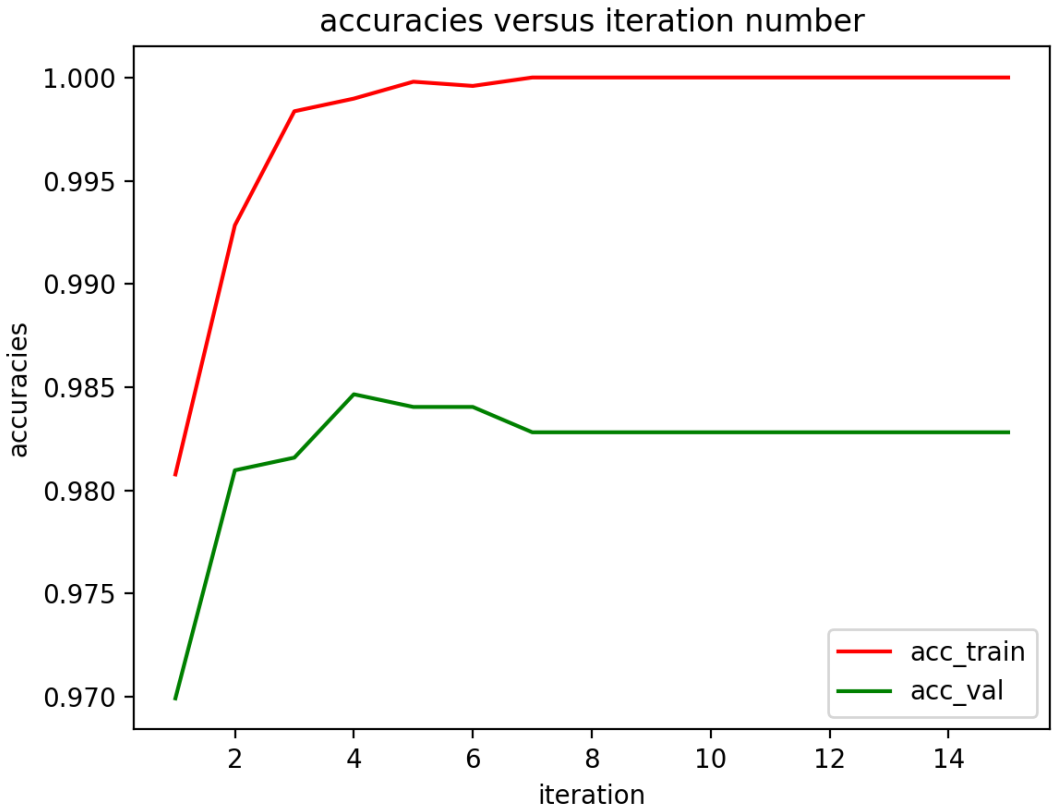
The best validation accuracy is 0.948435



[P = 2]

Iteration	Accuracy for the train	Accuracy for the validation
1	0.964403	0.945979
2	0.982406	0.973603
3	0.992635	0.979128
4	0.989157	0.976059
5	0.996727	0.980970
6	0.996727	0.979742
7	0.998363	0.979742
8	0.999591	0.983425
9	0.997750	0.981584
10	0.999591	0.981584
11	0.998773	0.978514
12	0.998773	0.979128
13	0.999591	0.979742
14	0.998363	0.980356
15	0.993249	0.972376

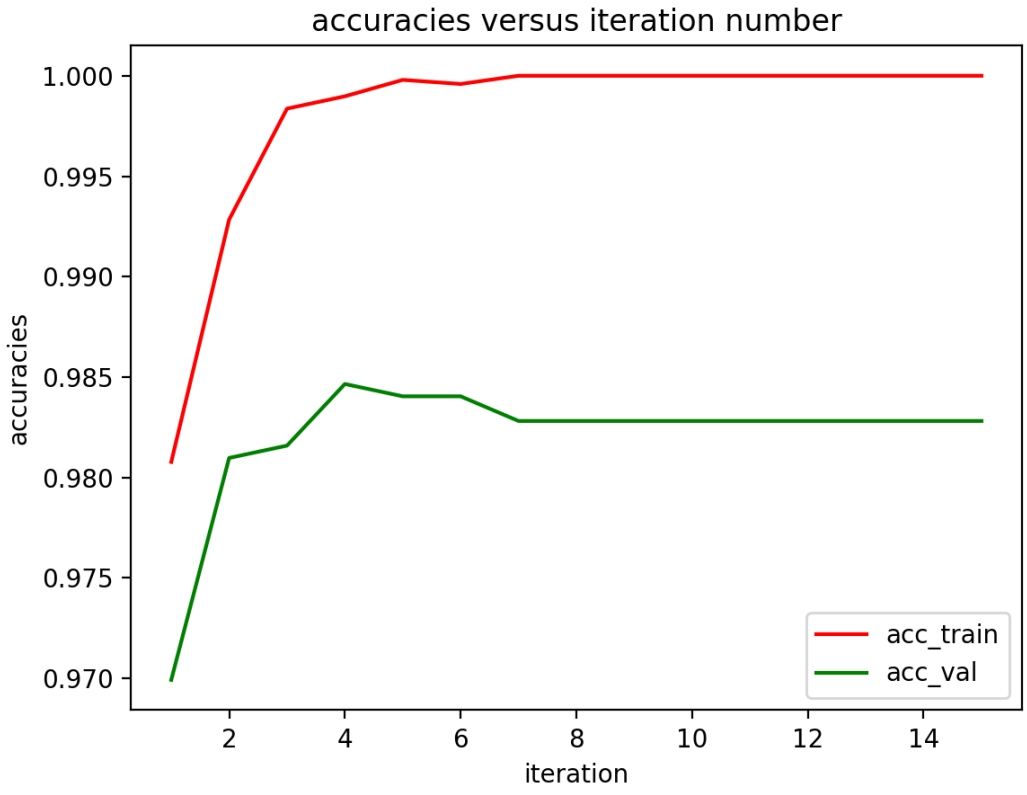
The best validation accuracy is 0.983425



[P = 3]

Iteration	Accuracy for the train	Accuracy for the validation
1	0.980769	0.969920
2	0.992840	0.980970
3	0.998363	0.981584
4	0.998977	0.984653
5	0.999795	0.984039
6	0.999591	0.984039
7	1.000000	0.982812
8	1.000000	0.982812
9	1.000000	0.982812
10	1.000000	0.982812
11	1.000000	0.982812
12	1.000000	0.982812
13	1.000000	0.982812
14	1.000000	0.982812
15	1.000000	0.982812

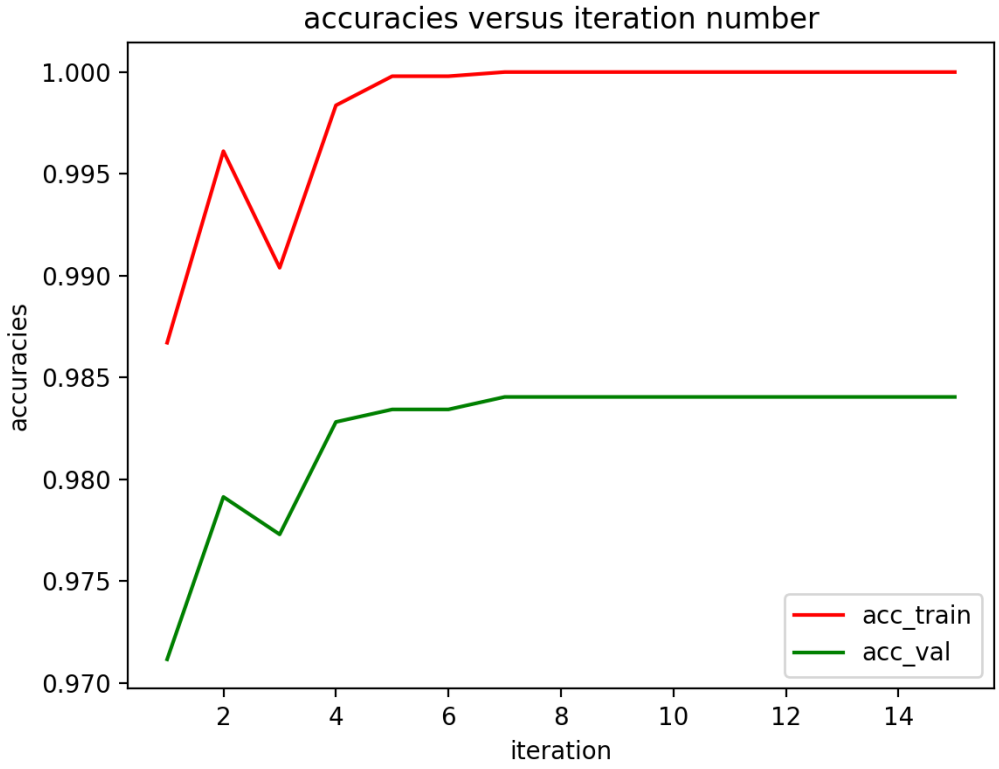
The best validation accuracy is 0.984653



[P = 4]

Iteration	Accuracy for the train	Accuracy for the validation
1	0.986702	0.971148
2	0.996113	0.979128
3	0.990385	0.977287
4	0.998363	0.982812
5	0.999795	0.983425
6	0.999795	0.983425
7	1.000000	0.984039
8	1.000000	0.984039
9	1.000000	0.984039
10	1.000000	0.984039
11	1.000000	0.984039
12	1.000000	0.984039
13	1.000000	0.984039
14	1.000000	0.984039
15	1.000000	0.984039

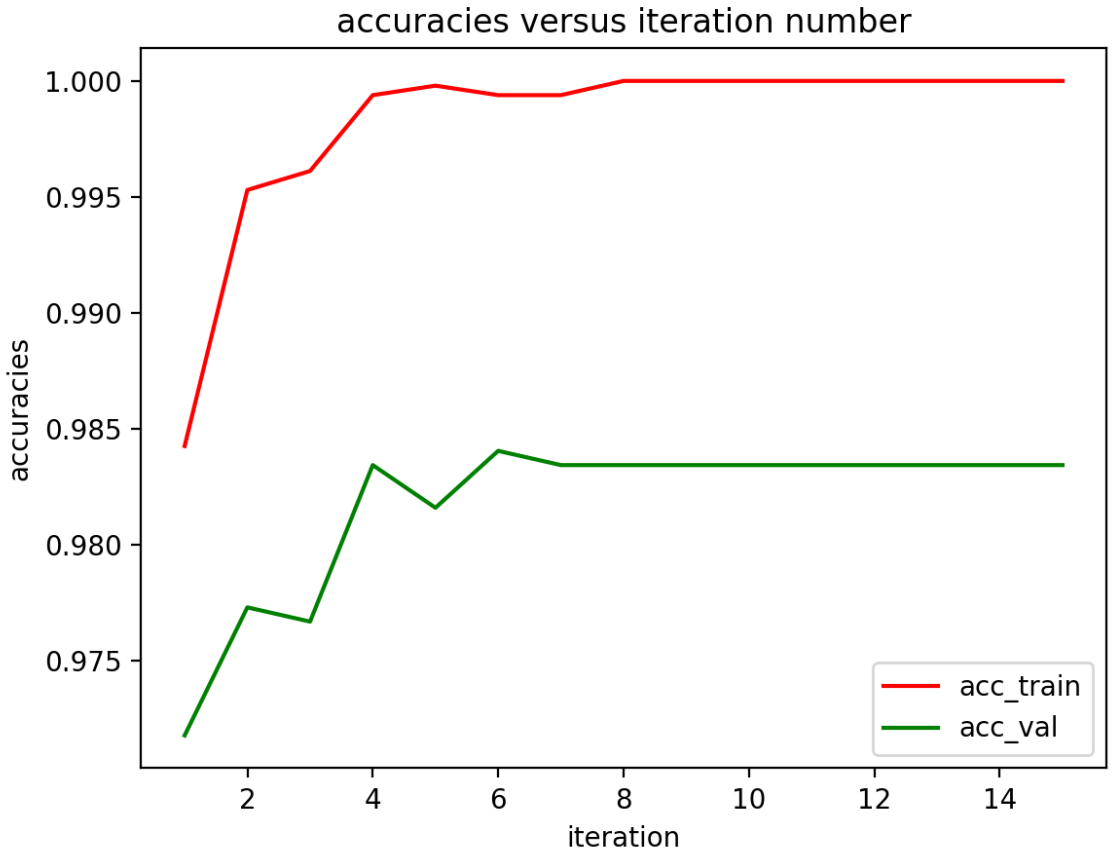
The best validation accuracy is 0.984039



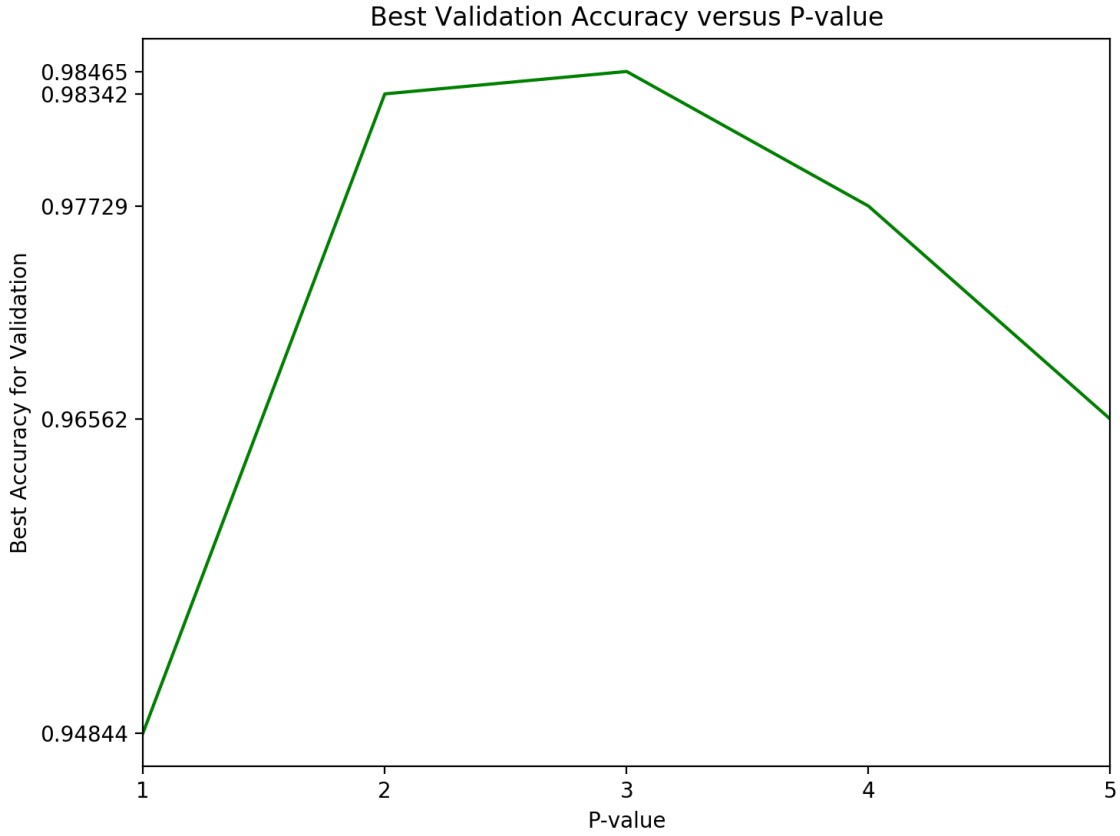
[P = 5]

Iteration	Accuracy for the train	Accuracy for the validation
1	0.984247	0.971762
2	0.995295	0.977287
3	0.996113	0.976673
4	0.999386	0.983425
5	0.999795	0.981584
6	0.999386	0.984039
7	0.999386	0.983425
8	1.000000	0.983425
9	1.000000	0.983425
10	1.000000	0.983425
11	1.000000	0.983425
12	1.000000	0.983425
13	1.000000	0.983425
14	1.000000	0.983425
15	1.000000	0.983425

The best validation accuracy is 0.984039



(a. 2)



From the above data, the P value affects the speed and stable to the predict accuracy. When we increase the P value, the accuracy will be fast even close to the 1(P value = 2, 3). However, if the P value keeps increasing, the accuracy will fall down quickly.

(b.)

The Best Value P = 3 and the prediction for the test set is in the "kplabel.csv"