

CS534 MACHINE LEARNING

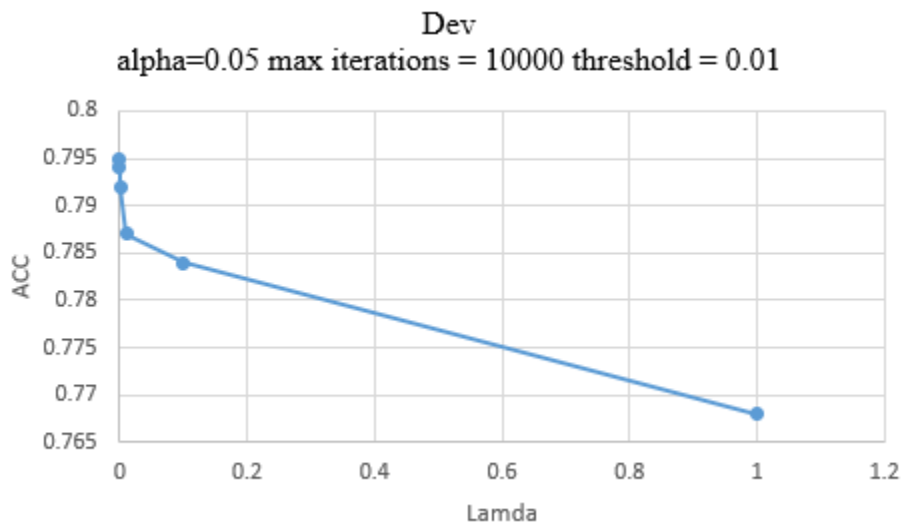
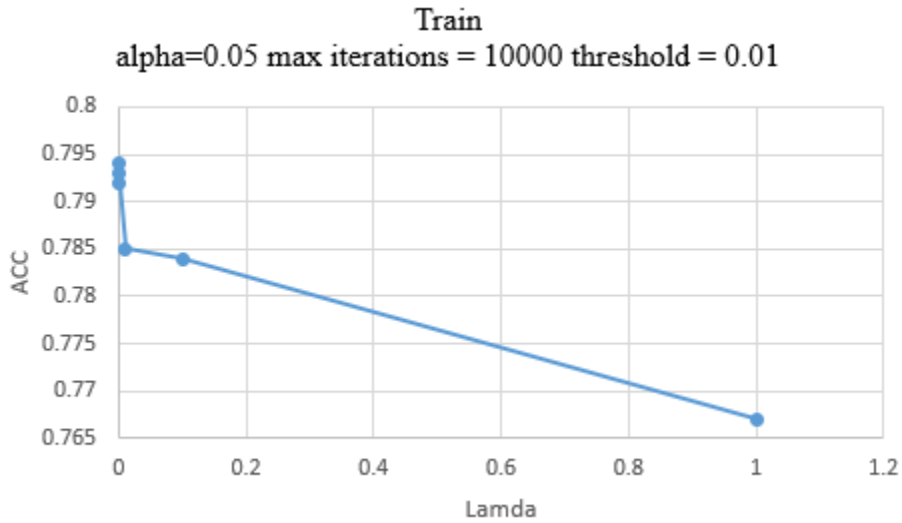
Da Lin 933334812

lind2@oregonstate.edu

lamda	accD2	accT2	accD1	accT1
1	0.498	0.5	0.768	0.767
0.1	0.784	0.785	0.784	0.784
0.01	0.784	0.785	0.787	0.785
0.001	0.786	0.785	0.792	0.792
0.0001	0.789	0.788	0.794	0.793
0.00001	0.79	0.789	0.795	0.794

Here is the overall result I use in below.

Part1:



b.

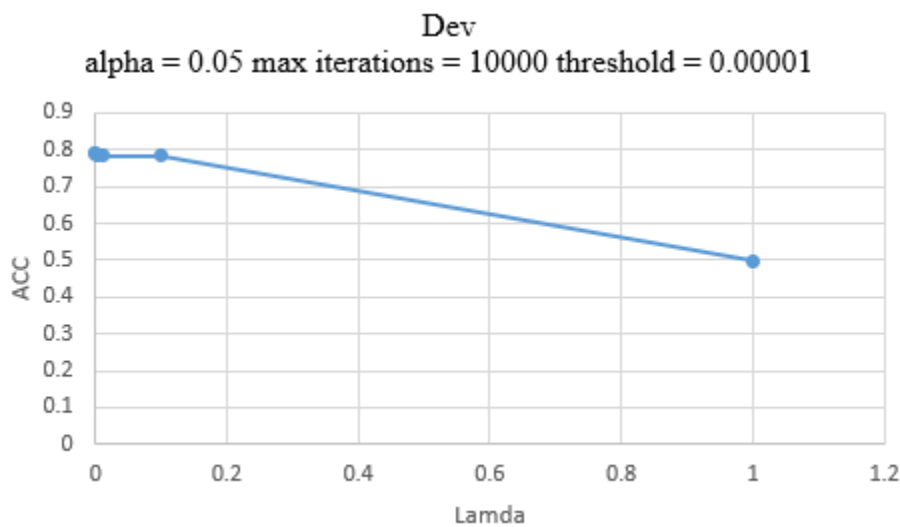
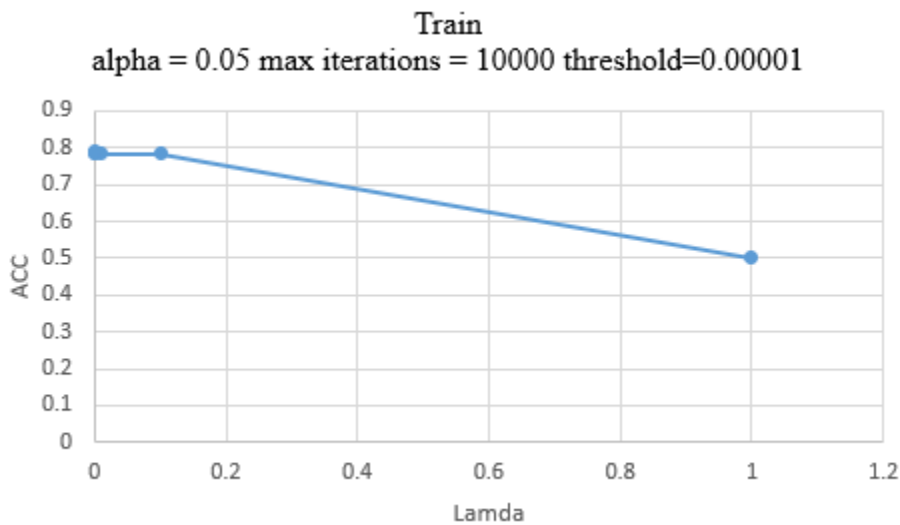
When lamda increased, the accuracy goes down, the reason is larger lamda train lesser data. The accuracy of Dev is a bit higher than Train, although they are similar. The best lamda is 0.00001.

c.

The best model we use is lamda = 0.00001. Top five is “Vehicle_Damage”, “Region_Code_28”, “Region_Code_41”, “Policy_Sales_Channel_26”, and “Policy_Sales_Channel_157”.

3 features have $w_j = 0$. When I am using different lamda, there are still 3 features have $w_j = 0$. So, when the lamda changed, numbers of features have $w_j = 0$ should be the same.

Part2:



b.

When lamda increased, the accuracy goes down, the reason is larger lamda train lesser data. However, the accuracy between lamda = 1 and lamda = 0.1 is change a lot, but when lamda goes enough small, the accuracy goes stable and did not change a lot. The accuracy of Dev is a bit higher than Train, although they are similar. The best lamda is 0.00001.

c.

The best model we use is lamda = 0.00001. Top five is “Vehicle_Damage”, “Region_Code_28”, “Policy_Sales_Channel_26”, “Policy_Sales_Channel_124”, “Policy_Sales_Channel_157”.

40 features have $w_j = 0$. When I increased lamda, numbers of features have $w_j = 0$ are also increase.

d.

Comparing L1 and L2, the funning time of L2 is higher than L1, but the accuracy of L2 is higher. When we use L1, the number of features have $w_j = 0$ is larger than using L2. In other words, the result of using L2 is smoother. The reason that using L1 cause the number of 0 changed is w_j base on the calculation of lamda, when lamda go smaller, the percentage that cause $w_j = 0$ is smaller.