Assignment 2

Guiliang Liu

October 22, 2016

1 Softmax for Multi-Class Classification

1. In the intersection point, the probabilities of the x belongs to Class1, Class2 and class 3 should equal to each other:

$$p(c_1|x) = p(c_2|x) = p(c_3|x) = \frac{1}{3}$$

2. For the points in the red line, their probabilities for the points belonging to the two classes which are divided by the red line are equal. For instance, for the points in the red line divides the class 1 and 2:

$$p(c_1|x) = p(c_2|x)$$

When we move along the red line, as mentioned before, the probabilities for the points belonging to the two classes which are divided by the red line are equal. However, the probability for the points belongings to the third class is reducing while the probability of belongings to the other two classes increase. In the example above, $p(c_3|x)$ will reduce and $p(c_1|x)$ and $p(c_2|x)$ will increase.

3. If we move far away from the intersection point and staying in the middle of one region, the probability for the points belonging to this region will increase and it will be larger than that of the remaining two regions(decrease as the point move). For example, if we stay in the region 1(Class 1):

$$p(c_1|x) > p(c_2|x)$$

and

$$p(c_1|x) > p(c_3|x)$$

2 Generalized Linear Models for Classification

1. The function $y_1(x)$ is Figure 1:

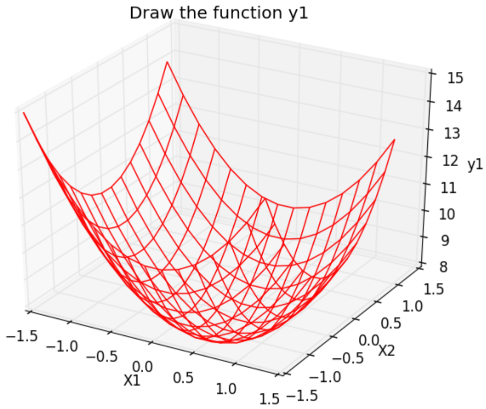


Figure 1: The function $y_1(x)$

2. The function $y_2(x)$ is Figure 2:

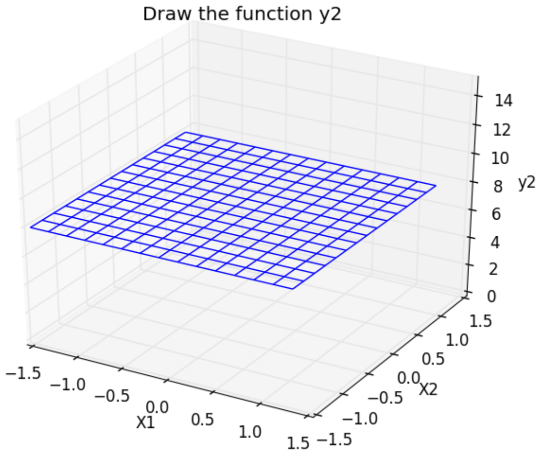
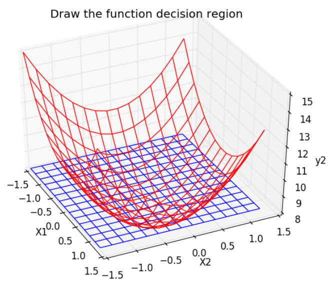
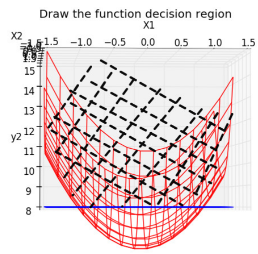


Figure 2: The function $y_2(x)$

3. The decision region is constructed by the intersection of $y_1(x)$ and $y_2(x)$, as showed below in the Figure 3(a), the y value of that region should be larger than 8, as the Figure 3(b) shows, the decision region is marked by the black dotted line. In the 2-D version, it should be a cycle: $y_1(x)^2 + y_2(x)^2 = 1$, the decision region is in Figure 4:



(a) intersection of $y_1(x)$ and $y_2(x)$



(b) the decision region

Figure 3: the plot of question 2.3

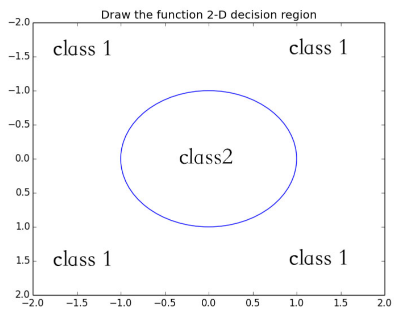
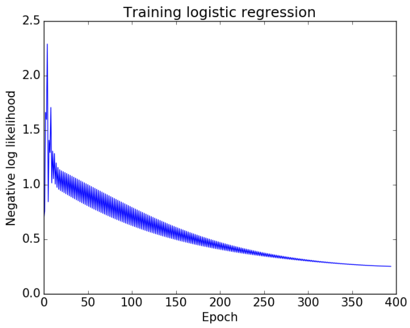


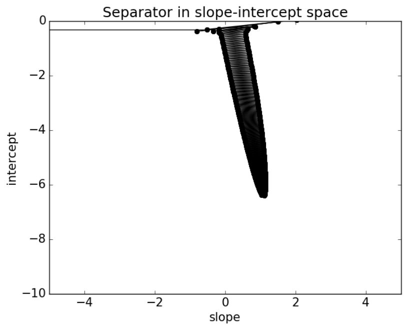
Figure 4: The 2-D decision region

3 Logistic Regression

1. The plot of separator path in slope-intercept space is showed in Figure 5(a) and the plot of neg. log likelihood over iterations is presented in Figure 5(b). The plot oscillating is caused by the learning rate and the direction of gradient descent. If the scale of the learning rate is 0.5, it will be too large and the gradient descent will jump too large in each iterations. Meanwhile, as the direction is not absolutely pointing to the bottom so it will induce lots of oscillation.



(a) the plot of neg. log likelihood over iterations



(b) the plot of slope-intercept space

Figure 5: question 3.1 The plot of separator path in slope-intercept space and the plot of neg. log likelihood over iterations

2. The plot comparing negative log-likelihood versus iteration for the different learning rates is presented in Figure 6. From Figure 6 we could conclude that as the learning rate η grows, the speed of convergence is improved but the oscillation will increase as well. But as the η become smaller, the regression will become more smooth but the speed will be reduced.

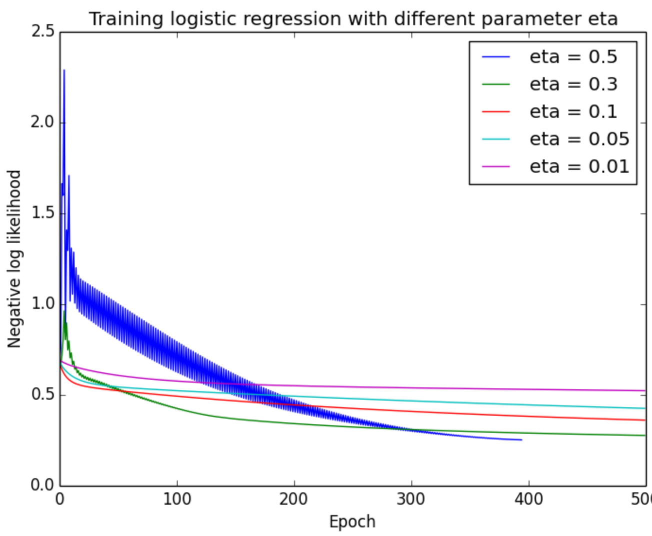


Figure 6: The plot comparing negative log-likelihood versus iteration for the different learning rates

3. The new plots of comparing negative log-likelihood versus iteration using stochastic gradient descent is showed in Figure 7. (I added 10^{-5} to all log inputs to avoid $log(0)$). Stochastic gradient descent is not always faster than the gradient descent, it relies on the scale of learning rate η. If η is large, the stochastic gradient descent will be slower but when the η is small, the stochastic gradient descent will be faster.

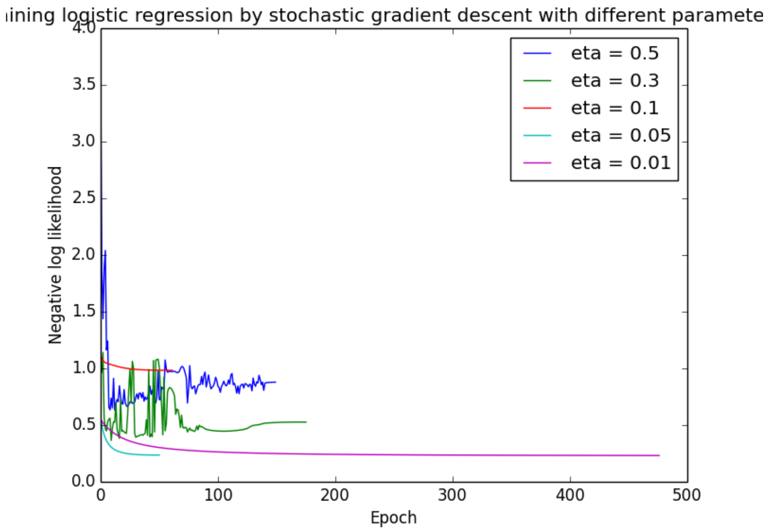


Figure 7: The stochastic gradient descent version of the plot comparing negative log-likelihood versus iteration for the different learning rates

4. The new plots of Figures 2 and 3 using IRLS is showed in Figure 8 and 9 if you add the constraint(plt.axis([-5, 5, -10, 0])), the image should be 9(a), if you delete the axis constrain , the image should be 9(b)

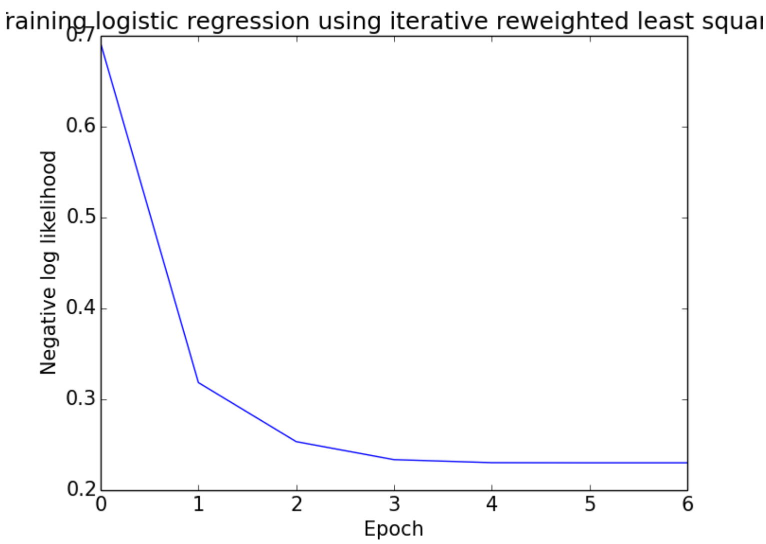
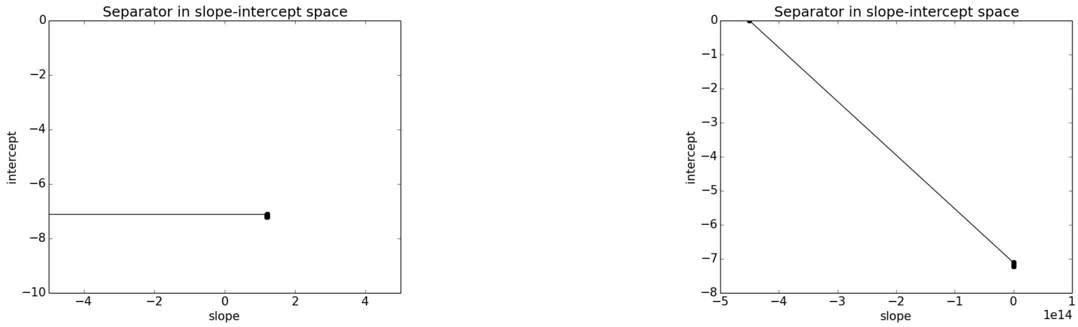


Figure 8: the plot of neg. log likelihood over iterations with IRLS



(a) the plot of slope-intercept space with IRLS with axis constrain

(b) the plot of slope-intercept space with IRLS without axis constrain

Figure 9: question 3.4 The plot of slope-intercept space