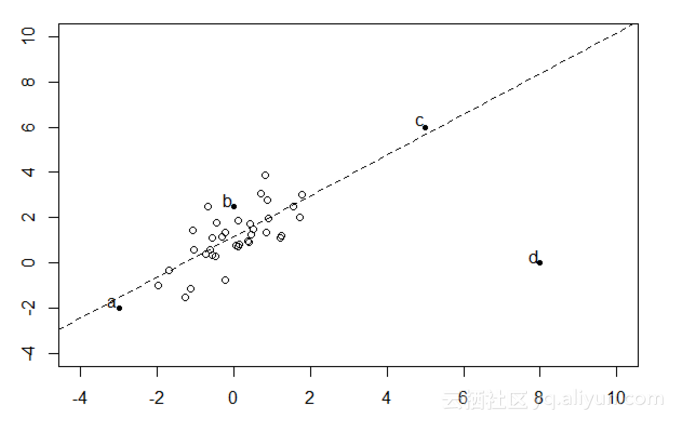
Writing exercise 2 in machine learning

Kyrie

崔颖琦

1730004002

1. Given the following data points, which point will affect the regression line (the dotted line) most if we remove it? Explain why?

A.a

B.b

C.c

D.d

Answer: C. Point c will affect the regression line most if we remove it. Because the point C is the closest point to the straight line of the four points. The parameters of the line is depend on the cost function and minimum the cost fucnction. Now point c is the closest point, it means that the the cost of point C is smallest, if point C move to another place, the cost will change a lot. So it will find the difference parameters to fit a line that minimum the cost function.

2. If we use one-vs-all strategy to a three class classification problem with three classes: -1, 0, 1, how many classifiers shall we train? What are they?

Answer: Three classifiers we should train. The first one is P(y=-1|x;). The second class is P(y=0|x;), and the third class is P(y=1|x;).

3. Describe the difference between linear regression and logistic regression. Please list at least three.

Answer: First, linear regression is used to solve prediction problems and logistic regression is used to solve classification.

Second, the variances of linear regression are continuous and variances of logistic regression are discrete.

Third, linear regression fit a straight line, the relationship of independent and dependent variables is linear. But in logistic regression, independent and dependent variables can be non-linear.

Fourth, the cost function of linear regression is the square error function. The cost function of logistic regression is sigmoid function.

4. Explain why don’t we use the cost function of linear regession as the cost function in sigmoid regession.

Answer: If we use square error fuction to be the cost function of sigmoid regression, the cost function is non-convex, it’s easy to get local minimum point, so we can’t get the global minimum point by gradient descent . That’s the reason why we don’t use the cost function of linear regression as the cost function in sigmoid regression.

5. Explain why do we want to make the Ɵ values small in regularization and how.

Answer: Because the there are too many features in the fuction, and high-order terms make the model overfit, but the features is important and we don’t want to reduce the number of the features. If we make the parameters of the high-order terms become small, so the multiplication of the parameters and the high-order terms will be small, so it cam make the function like a line of a quadratic function or some simple function. So we want to make the theta values small in regulariztion.

We add at the end of the cost function. Because we want to minimize the cost function, if we add regularization at the end of cost function, it will make the parameters become smaller by gradient descent. In the process of gradient descent, each step the parameters will minus a constant, so the parameters will become smaller. That’s how the regularizaion make the parameters smaller.

6. The following picture describes the decision boudary of logistic regression classification model with the same training data.

6.1 Which decision boundary overfitted the traning data? Why?

Answer: Graph C overfit the trainning data. Because the hypothesis fitted the training data too well, but it will fail to generalize to new examples.

6.2 If we get the decision boundaries after normalization with different regulalization parameter λ only, which one correponds to the largest regulalization parameter λ? Why?

Answer: A. Because If the regulalization parameter is very large, the parameters of the hypothesis will close to 0, except . So the hypothesis will like a constant straight line. Because are close to 0, the hypothesis will approximately equal to . So graph A corresponds to the largest regularization parameter.

