## Homework1: Build a classifier for iris type classification

姓名：王曌 学号：18210980068 专业：应用统计

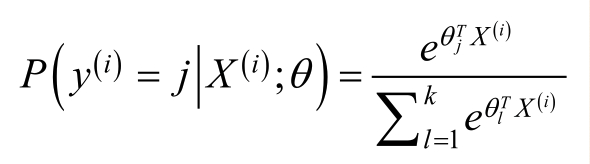
### I. Environment

★ Python 3.7.0

### II. Basic Theory

Softmax regression can be seen as an extension of logistic regression. In softmax regression, the output 'y' can take on multiple values.

1. The probability P(y=j | X) is shown as following.

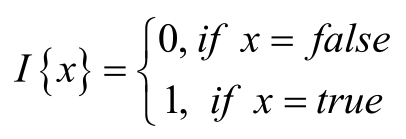


def prob(I, X, theta):

★ Calculating numerator

★ Calculating denominator

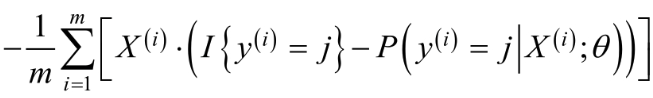
(2) The indicator function can be expressed as this.



def I(x, y):

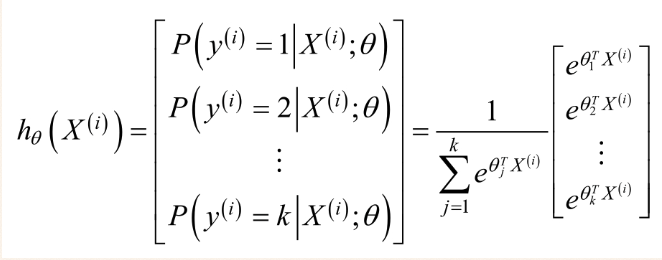
* Output 0 or 1

(3) The derivative of the gradient in each iteration for softmax regression:



def grad(j, theta)

(4) Calculate the probability of each kind.



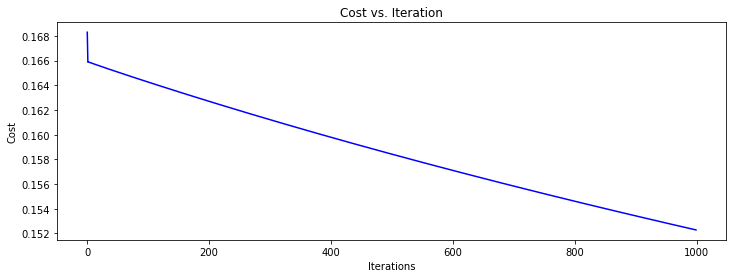
def h(x)

### III. Results

I use three fold cross validation for learning rate = 0.1, 0.01, 0.001.Code in ‘CrossValidation.py’

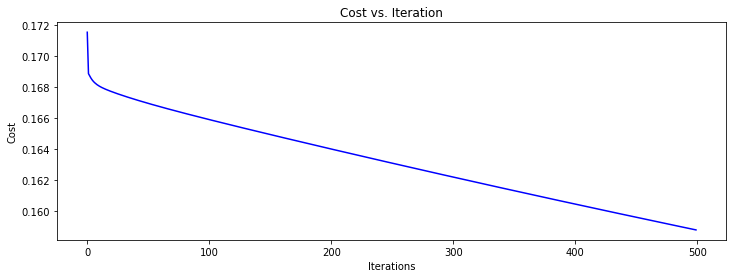
Consequently, when learning rate is 0.01, the model has highest accuracy

#### The first fold



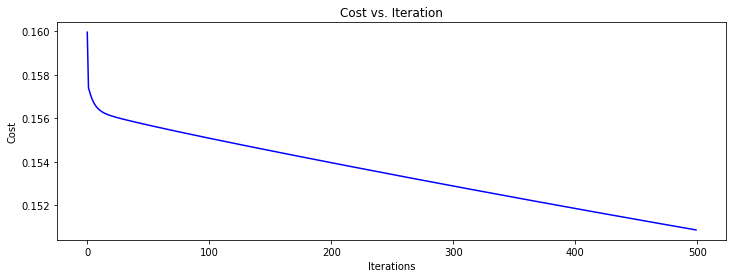
**Accuracy is 0.9714**

#### The second fold



**Accuracy is 0.9428**

#### The third fold



**Accuracy is 1**

**In N-fold Cross Validation, I can obtain the average accuracy is**

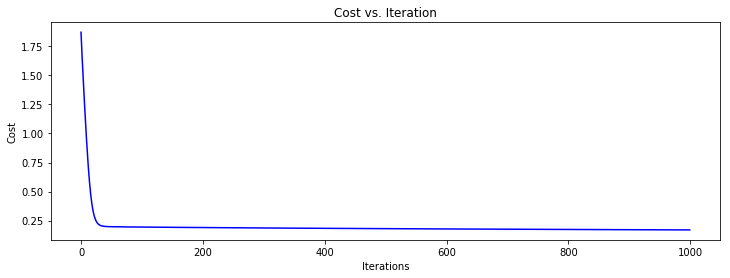
**(0.9714+0.9428+1)/3 = 0.9714**

|  |  |  |  |
| --- | --- | --- | --- |
| **Learning rate** | 0.1 | 0.01 | 0.001 |
| **Accuracy** | 0.94283 | 0.9714 | 0.74286 |

Finally I choose the hyperparameter lr = 0.01, iteration = 1000.The overall code in ‘SoftmaxRegression.py’

The Final **Accuracy is 0.9556**

Cost vs Iteration



### IV. Reference

1. [Softmax Regression算法实践](https://www.cnblogs.com/wanshuai/p/9106413.html)
2. [Softmax Regression模型](https://www.cnblogs.com/wanshuai/p/9104518.html)
3. [Pandas中DataFrame的分组/分割/合并](https://blog.csdn.net/shu15121856/article/details/84035501)