Homework1: Build a classifier for iris type classification

姓名：王嬿　　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.

$$P\left(y^{(i)}=j\middle|X^{(i)};\theta\right)=\frac{e^{\theta_j^T X^{(i)}}}{\sum_{l=1}^{k} e^{\theta_l^T X^{(i)}}}$$

def prob(I, X, theta):

★ Calculating numerator

★ Calculating denominator

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

$$I\{x\}=\begin{cases}0, if\ \ x=false\\ 1,\ \ if\ \ x=true\end{cases}$$

def I(x, y):

★ Output 0 or 1

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

$$-\frac{1}{m}\sum_{i=1}^{m}\left[X^{(i)}\cdot\left(I\left\{y^{(i)}=j\right\}-P\left(y^{(i)}=j\middle|X^{(i)};\theta\right)\right)\right]$$

def grad(j, theta)

(4) Calculate the probability of each kind.

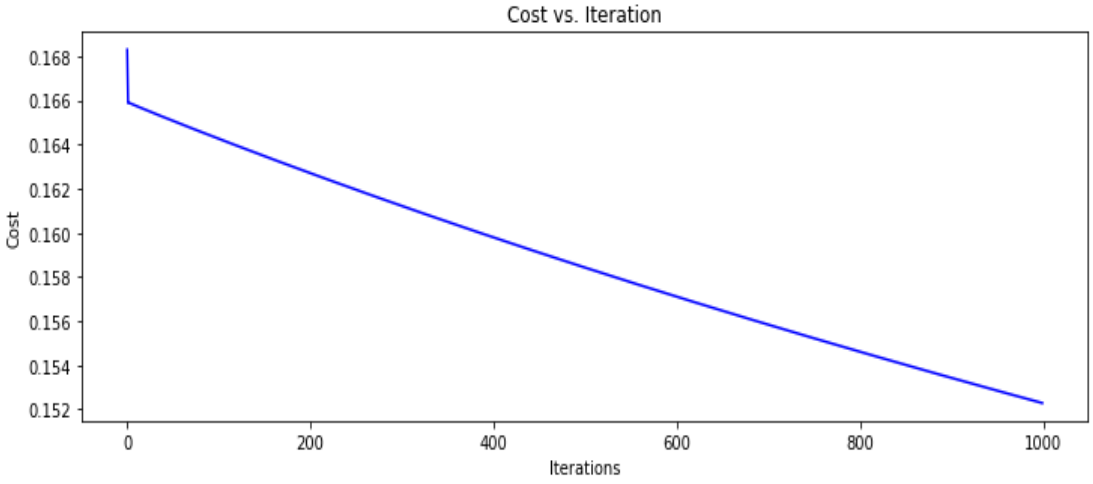
$$h_\theta\left(X^{(i)}\right)=\begin{bmatrix} P\left(y^{(i)}=1\middle|X^{(i)};\theta\right) \\ P\left(y^{(i)}=2\middle|X^{(i)};\theta\right) \\ \vdots \\ P\left(y^{(i)}=k\middle|X^{(i)};\theta\right) \end{bmatrix}=\frac{1}{\sum_{j=1}^{k} e^{\theta_j^T X^{(i)}}}\begin{bmatrix} e^{\theta_1^T X^{(i)}} \\ e^{\theta_2^T X^{(i)}} \\ \vdots \\ e^{\theta_k^T X^{(i)}} \end{bmatrix}$$

def h(x)

III. Results

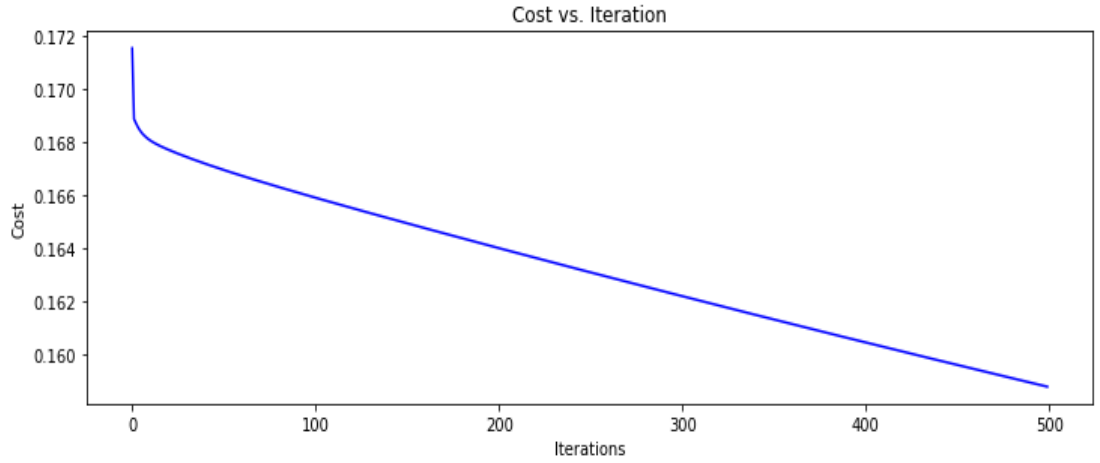
I use three fold cross validation for learning rate = 0.1, 0.01, 0.001.Code in 'CrossValidation.py' Consequently, when learning rate is 0.01, the model has highest accuracy

The first fold



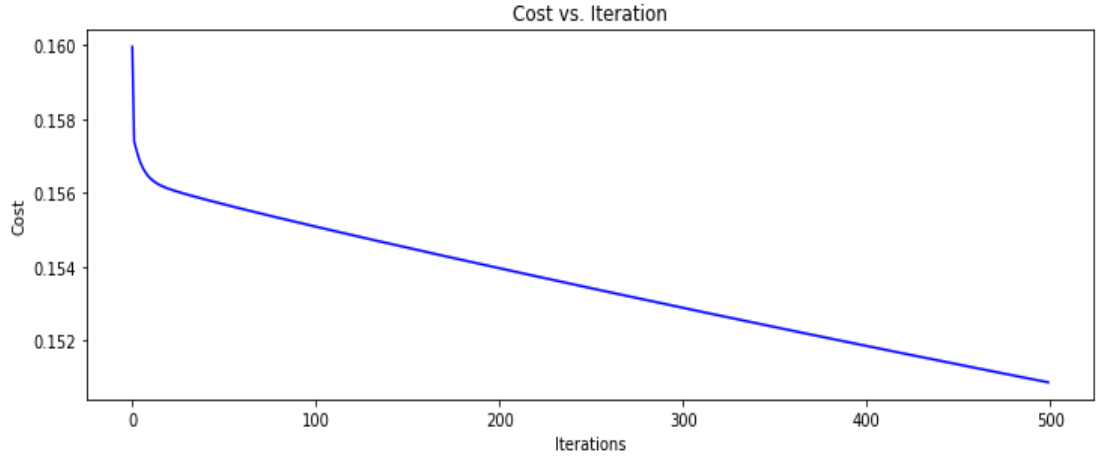
Accuracy is 0.9714

The second fold



Accuracy is 0.9428

The third fold



Accuracy is 1

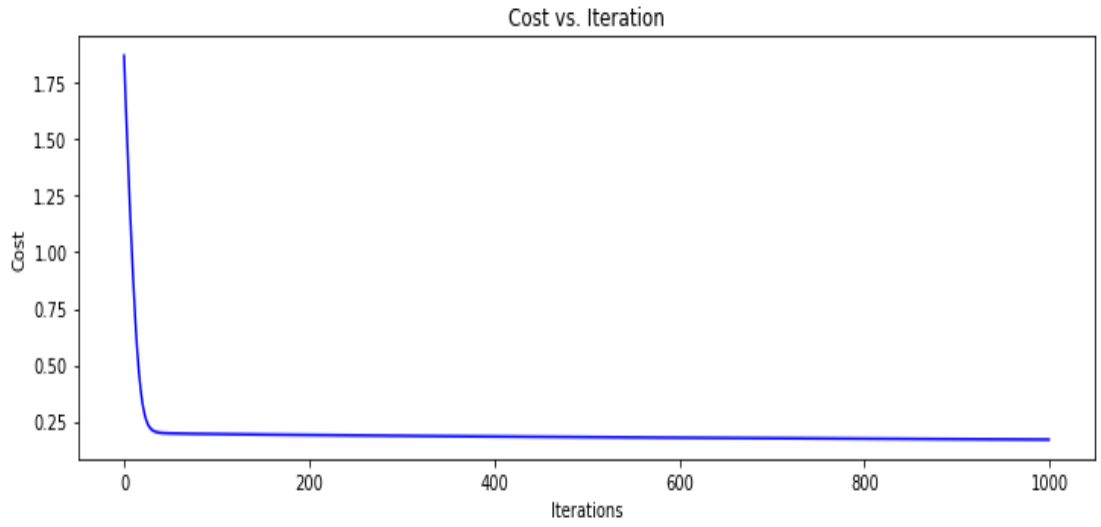
In N-fold Cross Validation, I can obtain the average accuracy is (0.9714+0.9428+1)/3 = 0.9714

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
2. Softmax Regression 模型
3. Pandas 中 DataFrame 的分组/分割/合并