**8.2**

代码：

*# 读取数据*

*import pandas as pd*

*file\_path = r"C:\Users\Mac\Desktop\过程\学业\本科\专业课\多元统计\LDA\data\wine.train.txt"*

*data = pd.read\_table(file\_path, header=None)*

*X = data.drop(columns=13)*

*y = data.iloc[:, -1]*

*# LDA及其结果展示*

*from sklearn.discriminant\_analysis import LinearDiscriminantAnalysis*

*lda = LinearDiscriminantAnalysis(solver='svd', n\_components=2)*

*X\_lda = lda.fit\_transform(X, y)*

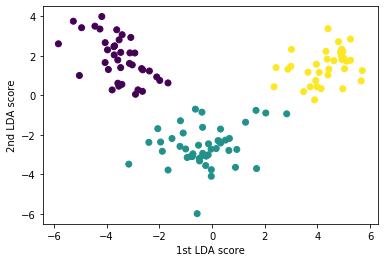
*# 绘图*

*import matplotlib.pyplot as plt*

*plt.scatter(X\_lda[:,0], X\_lda[:,1], c=y)*

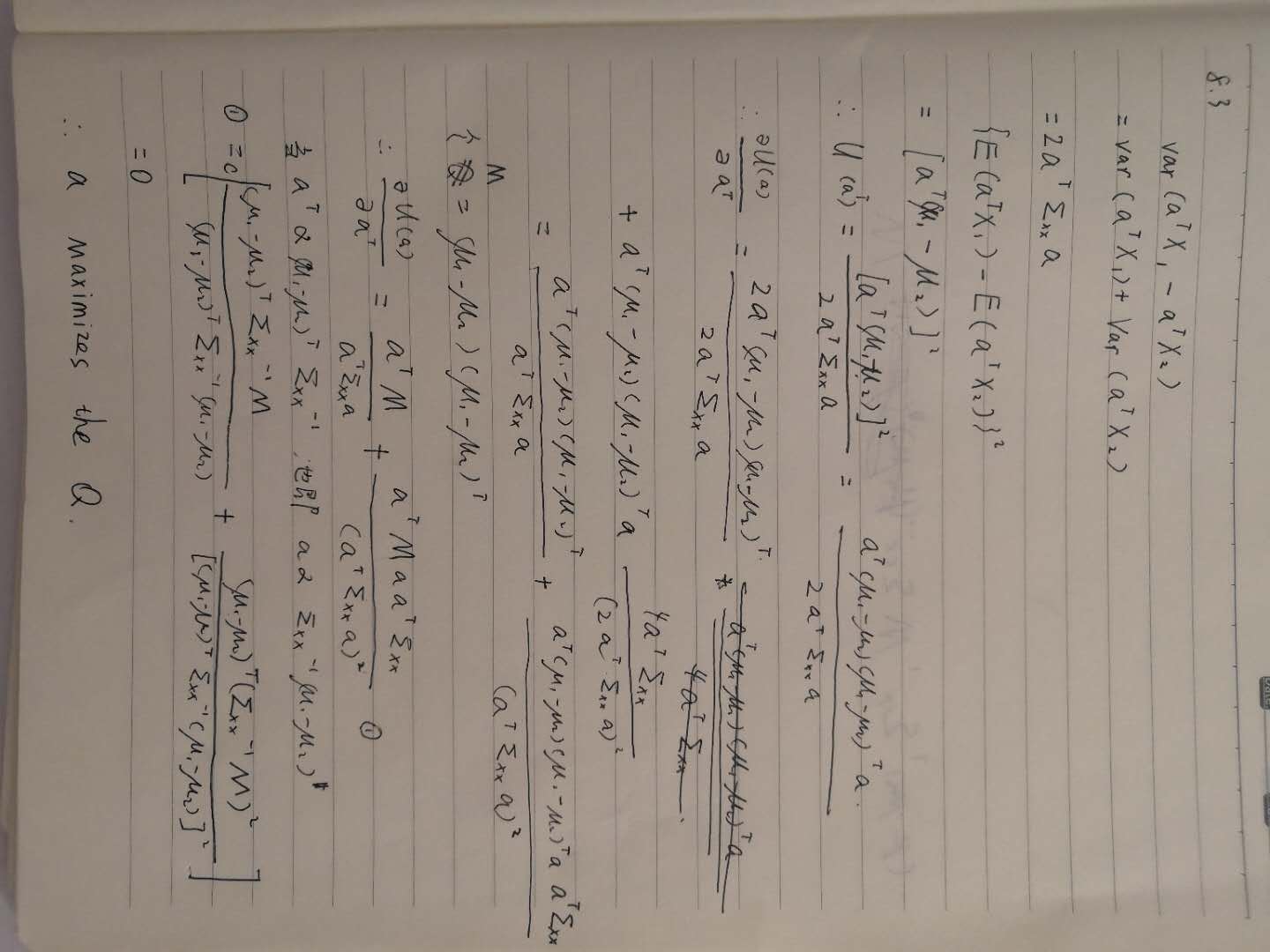
*plt.xlabel("1st LDA score")*

*plt.ylabel("2nd LDA score")*

**

可以看出，前二个LDA成分对模型的分类效果较好。

**8.3**



**8.6**

代码：

*# 读取数据（这个数据集有150个样本）*

*import seaborn as sns*

*data = sns.load\_dataset("iris")*

*X = data.drop(columns='species')*

*y = data['species']*

*# 数据转化（相除，取对数）*

*X['sepal\_shape'] = X['sepal\_length'] / X['sepal\_width']*

*X['petal\_shape'] = X['petal\_length'] / X['petal\_width']*

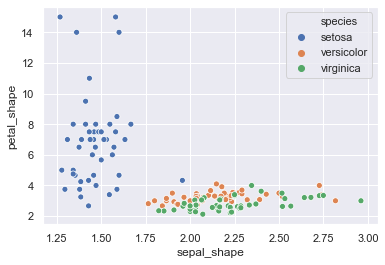
*import math*

*X\_tr = X[['sepal\_shape', 'petal\_shape']].applymap(lambda x: math.log(x))*

*# 绘制初始变量散点图*

*sns.set()*

*sns.scatterplot(X['sepal\_shape'], X['petal\_shape'], hue=y)*

**

*# LDA & QDA*

*from sklearn.discriminant\_analysis import LinearDiscriminantAnalysis*

*from sklearn.discriminant\_analysis import QuadraticDiscriminantAnalysis*

*from sklearn.model\_selection import cross\_val\_predict, LeaveOneOut*

*from sklearn.metrics import accuracy\_score*

*loo = LeaveOneOut()*

*# 输出LDA的cv误分类率*

*lda = LinearDiscriminantAnalysis(solver='svd', n\_components=2)*

*y\_lpred = cross\_val\_predict(lda, X, y, cv=loo)*

*print("LDA的cv误分类率为：", 1 - accuracy\_score(y, y\_lpred))*

*LDA的误分类率为： 0.026666666666666616*

*# 输出QDA的cv误分类率*

*qda = QuadraticDiscriminantAnalysis()*

*y\_qpred = cross\_val\_predict(qda, X, y, cv=loo)*

*print("QDA的cv误分类率为：", 1 - accuracy\_score(y, y\_qpred))*

*QDA的误分类率为： 0.040000000000000036*