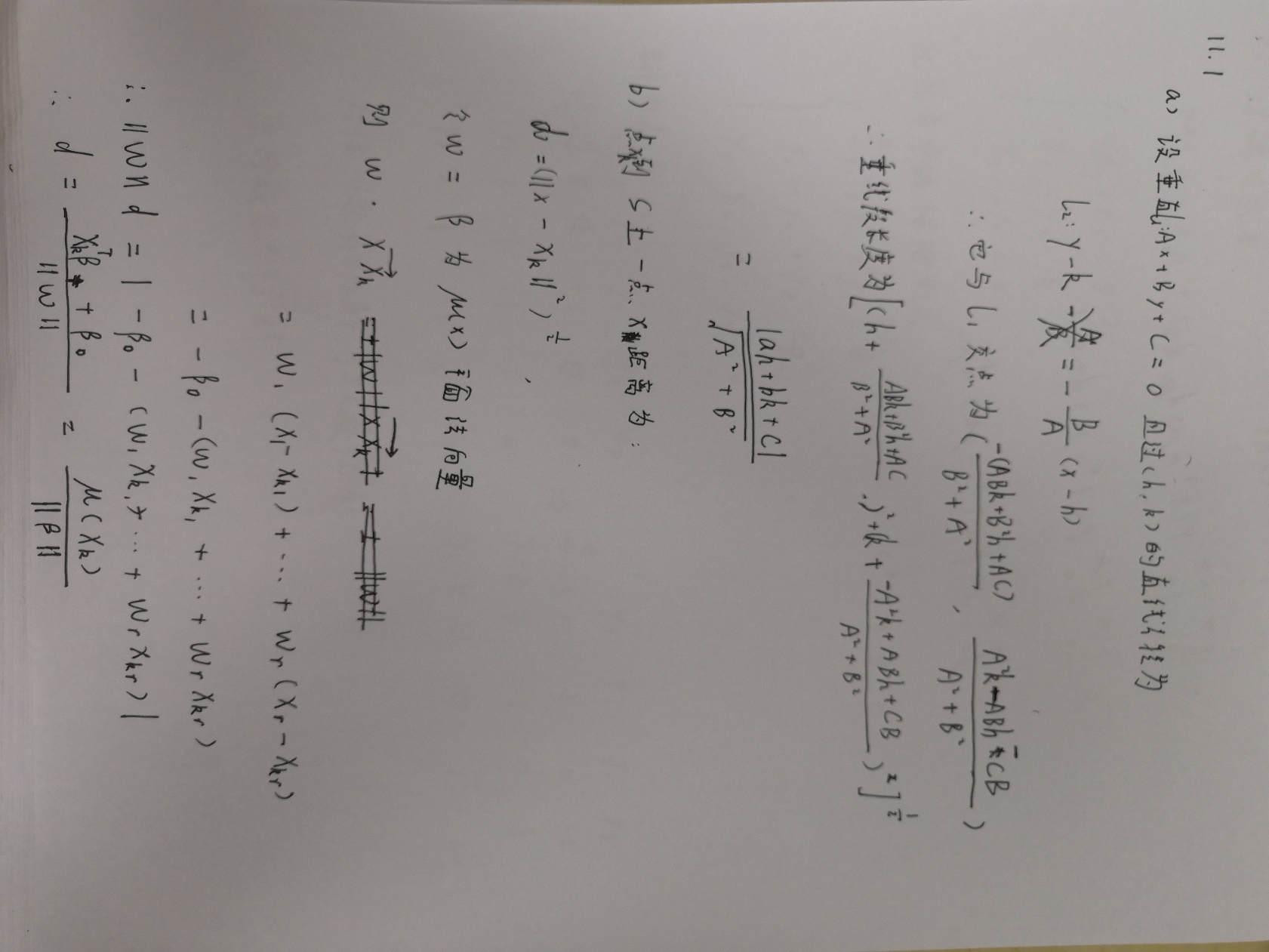
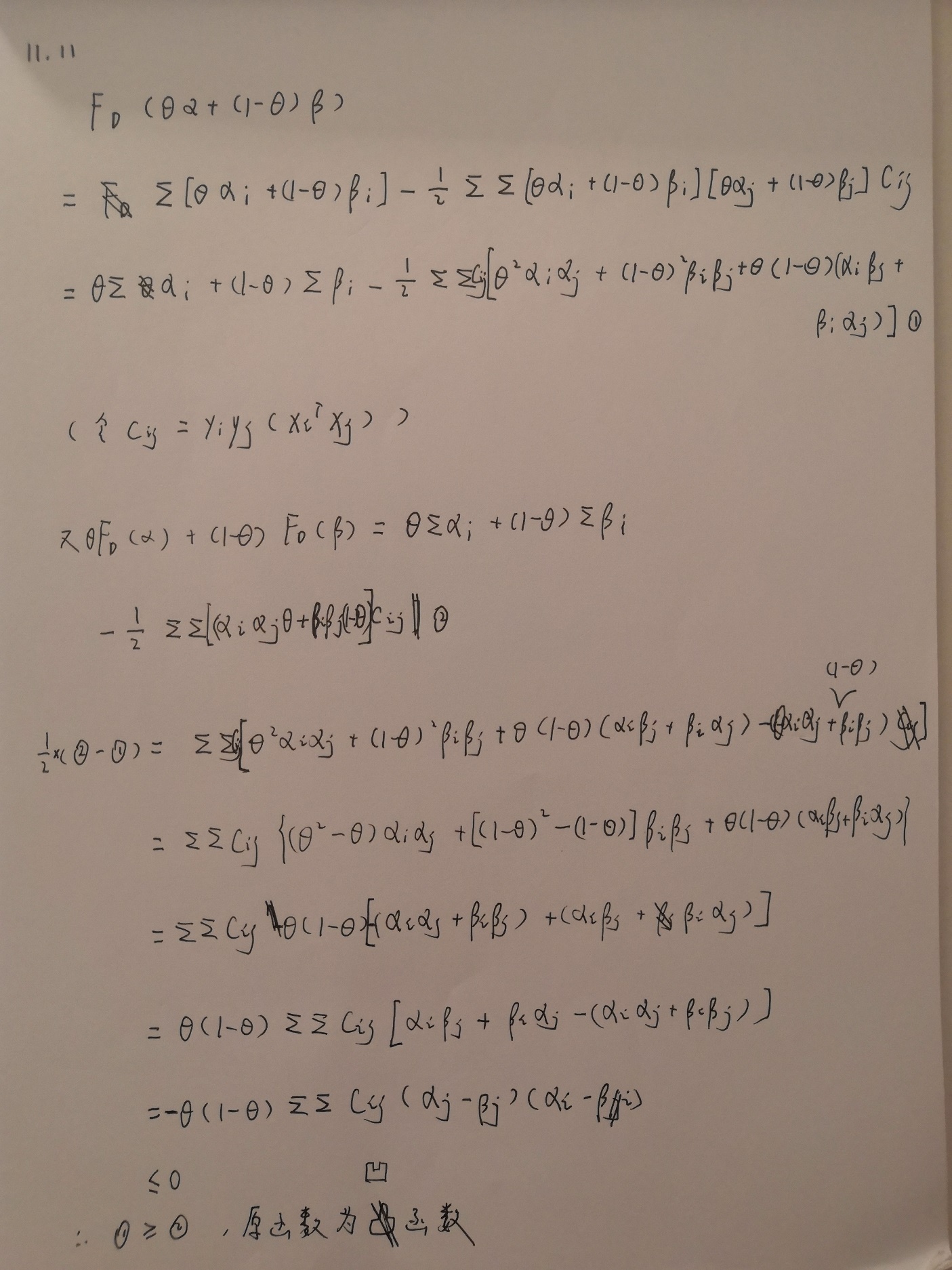
**11.1**

**11.11**

****

**11.12**

**代码如下：**

*## 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\wdbc.txt"*

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

*X = data.drop(columns=[0, 1]).values*

*y = data.loc[:, 1].values*

*## 网格搜索SVM的最佳参数，并得到对应的cv误判率*

*import numpy as np*

*from sklearn.svm import SVC*

*from sklearn.model\_selection import KFold, GridSearchCV*

*# 设置cv折数，定义参数表格*

*fold = KFold(n\_splits=10, random\_state=42)*

*param\_range = {'C': [1, 10, 50, 100, 500, 1000],*

*'gamma': np.logspace(-5, -1, 5)}*

*# 进行网格搜索*

*svm\_clf = SVC(kernel='rbf')*

*grid\_cv = GridSearchCV(svm\_clf, cv=fold, param\_grid=param\_range)*

*grid\_cv.fit(X, y)*

*# 得到所有组合的cv误判率和最佳参数组合*

*svm\_mcr = pd.DataFrame(*

*(1 - grid\_cv.cv\_results\_['mean\_test\_score']).reshape(6, -1),*

*columns=param\_range['gamma'], index=param\_range['C'])*

*print("高斯核SVM各参数组合的cv误判率表，行为C值，列为gamma值\n", svm\_mcr)*

*print("高斯核SVM最佳参数组合为：", grid\_cv.best\_params\_, "；对应的cv误判率为：", svm\_mcr.min().min())*

*## 得到LDA和TREE分类器的cv误判率*

*from sklearn.model\_selection import KFold*

*from sklearn.model\_selection import cross\_val\_score*

*fold = KFold(n\_splits=10, random\_state=42)*

*# LDA模型的cv误判率*

*from sklearn.discriminant\_analysis import LinearDiscriminantAnalysis*

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

*print("LDA的cv误判率为：", 1 - cross\_val\_score(lda\_clf, X, y, cv=fold).mean())*

*# Tree模型的cv误判率*

*from sklearn.tree import DecisionTreeClassifier*

*tree\_clf = DecisionTreeClassifier(criterion='gini', splitter='best', max\_depth=5)*

*print("决策树的cv误判率为：", 1 - cross\_val\_score(tree\_clf, X, y, cv=fold).mean())*

**输出如下**：

高斯核SVM各参数组合的cv误判率表，行为C值，列为gamma值

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 0.00001 | 0.0001 | 0.001 | 0.01 | 0.1 |
| 1 | 0.077329 | 0.066784 | 0.082601 | 0.369069 | 0.372583 |
| 10 | 0.065026 | 0.068541 | 0.093146 | 0.367311 | 0.372583 |
| 50 | 0.057996 | 0.059754 | 0.093146 | 0.367311 | 0.372583 |
| 100 | 0.052724 | 0.057996 | 0.093146 | 0.367311 | 0.372583 |
| 500 | 0.045694 | 0.070299 | 0.093146 | 0.367311 | 0.372583 |
| 1000 | 0.049209 | 0.070299 | 0.093146 | 0.367311 | 0.372583 |

高斯核SVM最佳参数组合为： {'C': 500, 'gamma': 1e-05} ；对应的cv误判率为： 0.04569420035149385

LDA的cv误判率为： 0.04044486215538856

决策树的cv误判率为： 0.06672932330827075