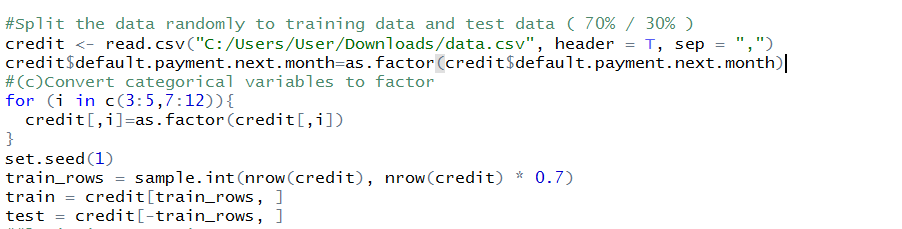
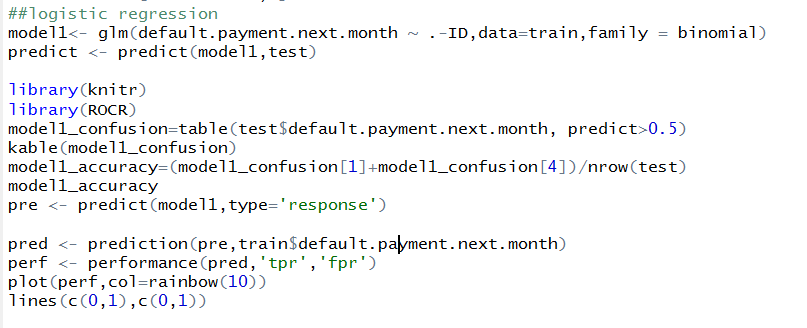
資料探勘 Homework2

0551287 土木所碩一 張為舜

1. Split the data randomly to training data and test data ( 70% / 30% ) 將資料切成70%的訓練資料，30%的測試資料



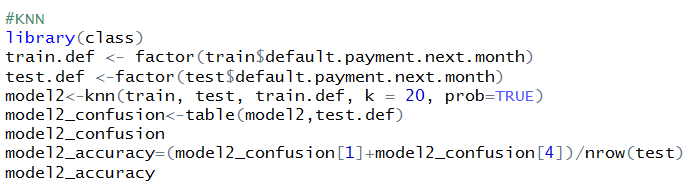
2. What is the **accuracy** of (1) Logistic Regression (2) k-Nearest Neighbors (3) Naive Bayes (4) Random Forest (5) SVM model in test data 用五種模型訓練，在測試資料的準確度分別是多少

**Logistic Regression:** 

**準確度: 81.2%**



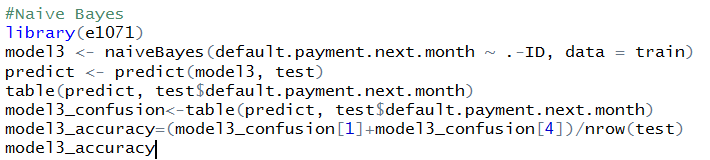
**k-Nearest Neighbors:**



**準確度:77.5%**



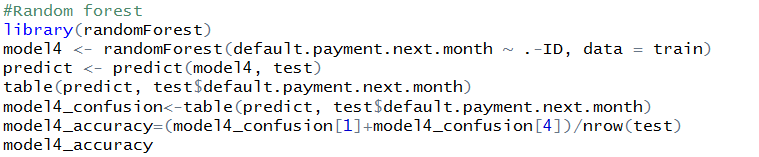
**Naive Bayes**



**準確度:70.3%**



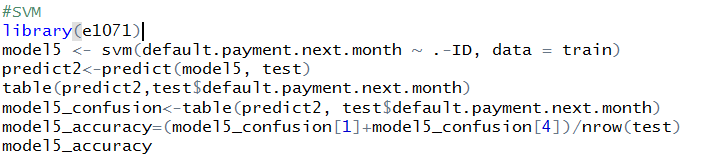
**Random Forest:**



**準確度:81.4%**



**SVM model in test data:**

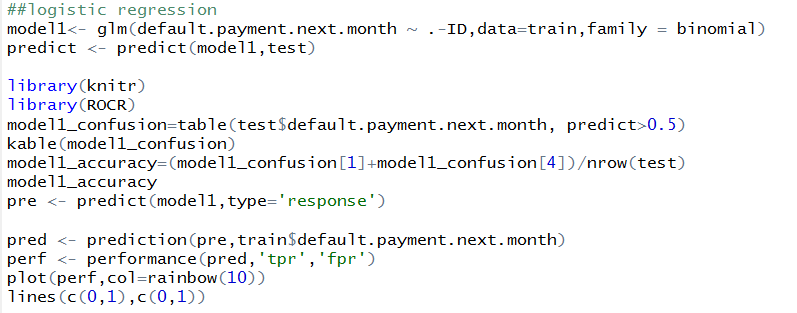


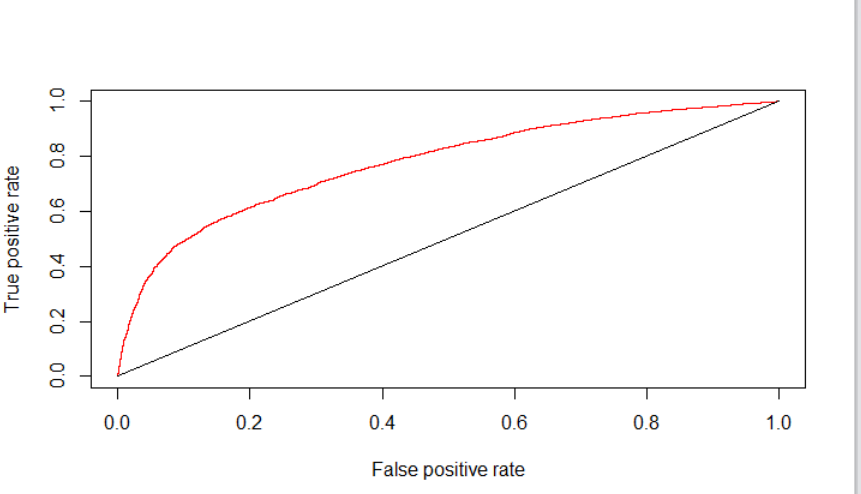
**準確度:81.6%**



**3. Draw the ROC curve in Logistic Regression**

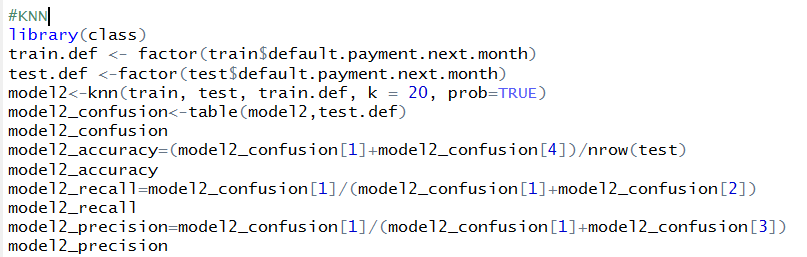
**畫出Logistic Regression的ROC curve**

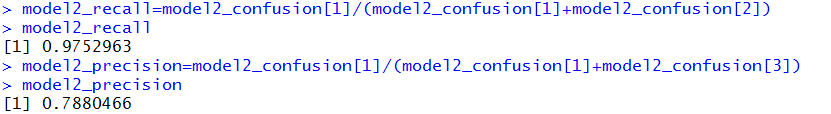




**4. Calculate the precision and recall in k-Nearest Neighbors**

**計算k-Nearest Neighbors的precision和recall**



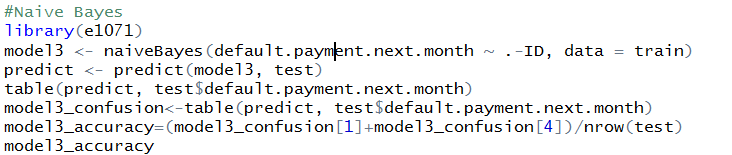


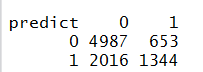
**Recall:97.5%**

**Precision:78.8%**

**5. Draw the Confusion Matrix of Naive Bayes**

**畫出Naive Bayes的Confusion Matrix**

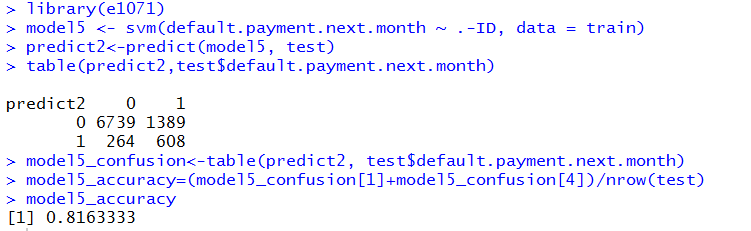




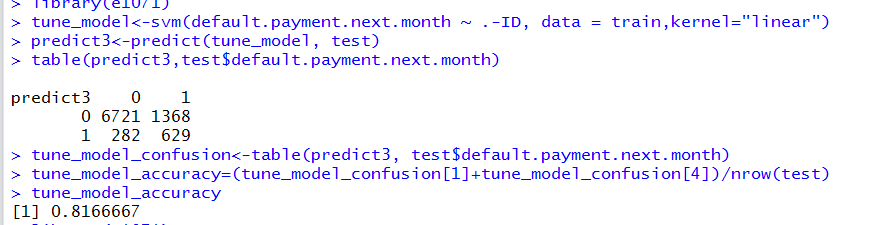
**6. What is the performance with different parameters in SVM**

**不同參數差異在SVM模型的表現**

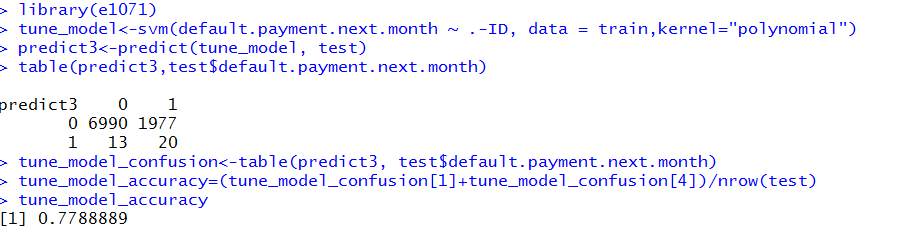
**使用預設值kernel=”radial” 準確度為81.63%**



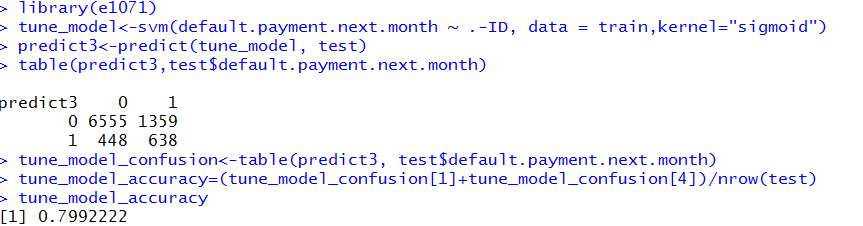
**使用kernel=”linear” 準確度為81.67%**



**使用kernel="polynomial" 準確度為77.89%**

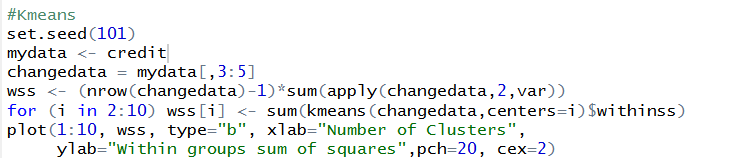


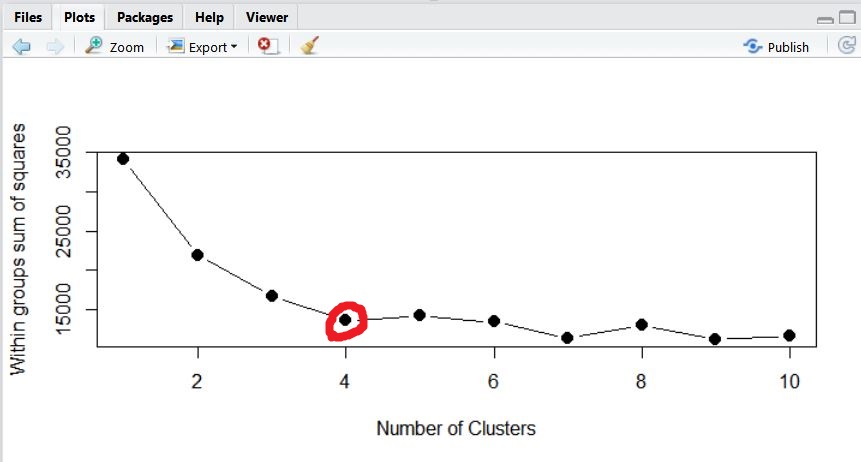
**使用kernel="sigmoid" 準確度為79.92%**



**結果為在kernel=”linear” 時 準確度最高**

**7. According to Gender, Education, and Marital status, how many kinds of customer should be divided into? ( Hint: use the elbow method with K-means) 你認為根據Gender , Education, and Marital status三個欄位，可以將所有客戶分成幾類? (提示:利用K-Means，elbow method)**

****

****

**A:透過elbow method可以得出透過Gender , Education, and Marital status三個欄位可以將所有客戶分為四類(彎曲點為4)**