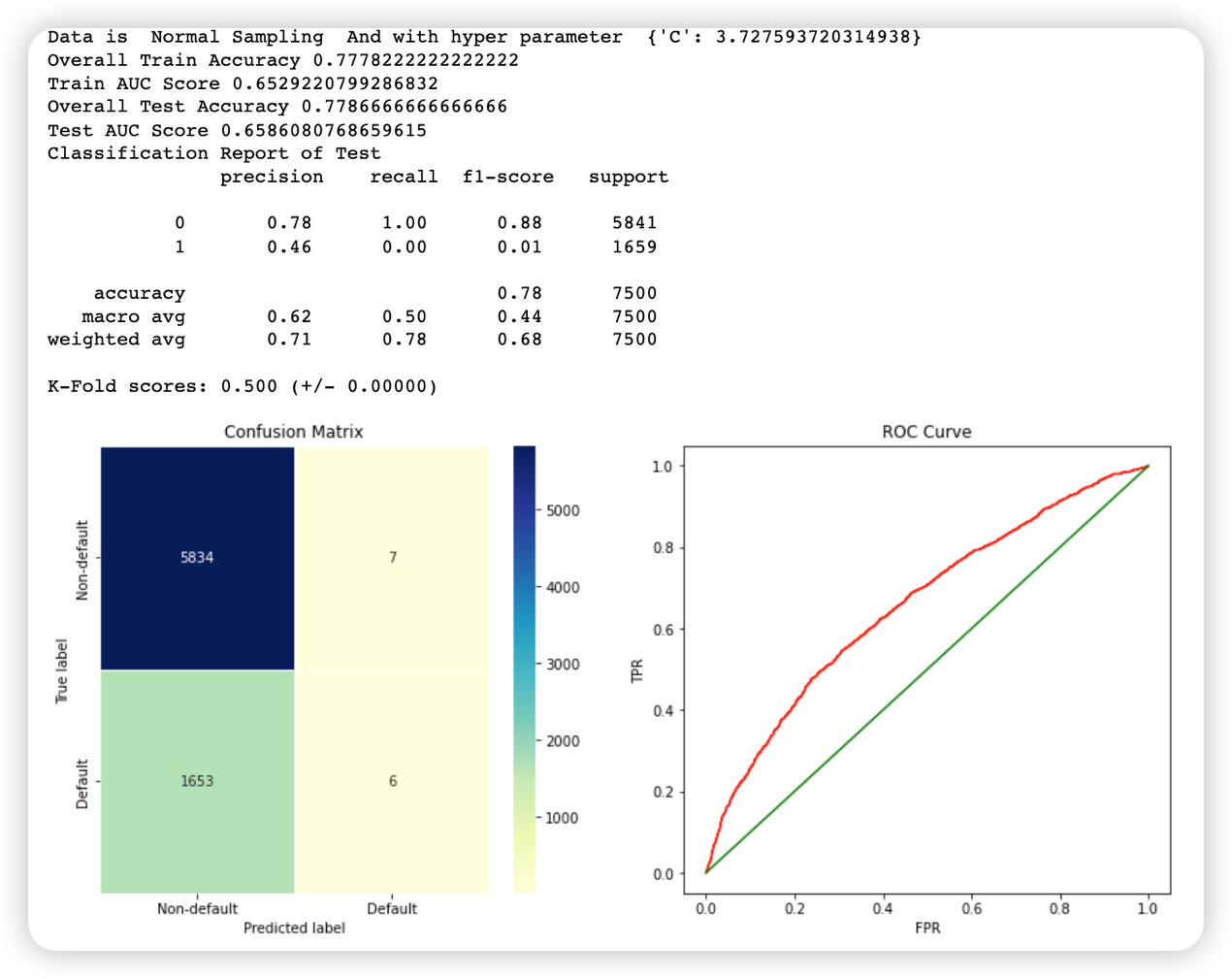
(1) Report the overall error rate and AUC of three models on the training and test sets and compare them.

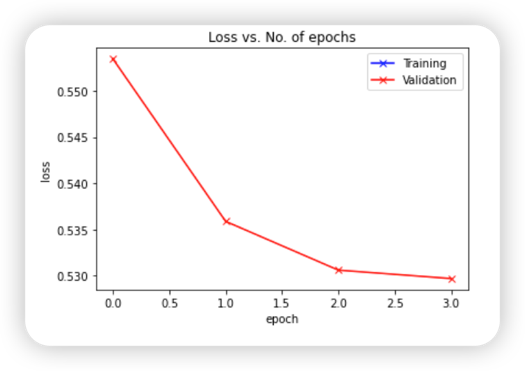
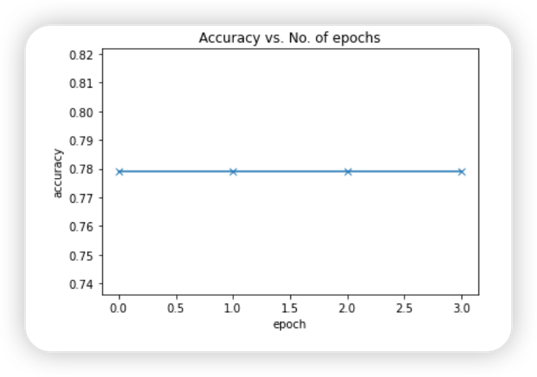
**Logistic Regression**



Accuracy is 78%, so the error rate is 22%.

AUC score is 0.65.

**FNN:**

****

In FNN model, the result as same as logistic regression.

**Classification Tree:**

Classification tree provides more accuracy than FNN or logistic regression with 17.9% error rates.

一張含有 桌 的圖片

自動產生的描述

一張含有 文字 的圖片

自動產生的描述一張含有 桌 的圖片

自動產生的描述

(2) Which variables are important for predicting defaults?

**Education and Repayment status(PAY\_X)** are important for predicting defaults.

一張含有 桌 的圖片

自動產生的描述

As value of PAY\_X increasing, the percentage of default will also be increasing, which demonstrate it have positive correlation to default.

一張含有 桌 的圖片

自動產生的描述

An group of education shows less default percentages than other groups.

Consider other variables, like SEX to default, MARRIAGE to default, related percentage are close to each data point.

一張含有 桌 的圖片

自動產生的描述一張含有 桌 的圖片

自動產生的描述

Age to default, related percentage follows a normal distribution. So these variables change, it wouldn’t affect a lot the result of default.

一張含有 桌 的圖片

自動產生的描述