Data Preparation

Download NVDA stock data from Yahoo Finance. Stock data includes two fields: date and adjusted stock price. The time span of the stock data is from Jan 2, 2016 to Dec 31, 2019, with a total of 1006. Perform logarithmic processing on adjusted stock price and then calculate logarithmic return.

$$LogR = Log(P_t - P_{t-1})$$

Form table 1, the highest adjusted price of NVDA in the past four years is 287.77 on Oct 1,2018 and the lowest price is 24.700 on Feb 8,2016. The highest log return is 0.261 and the lowest is -0.208.

Table 1. Data Description

	Min	1st Qu	Median	Mean	3rd Qu	Max
Adj Close	24.700	98.140	160.050	151.700	209.180	287.770
Log R	-0.208	-0.010	0.002	0.002	0.015	0.261

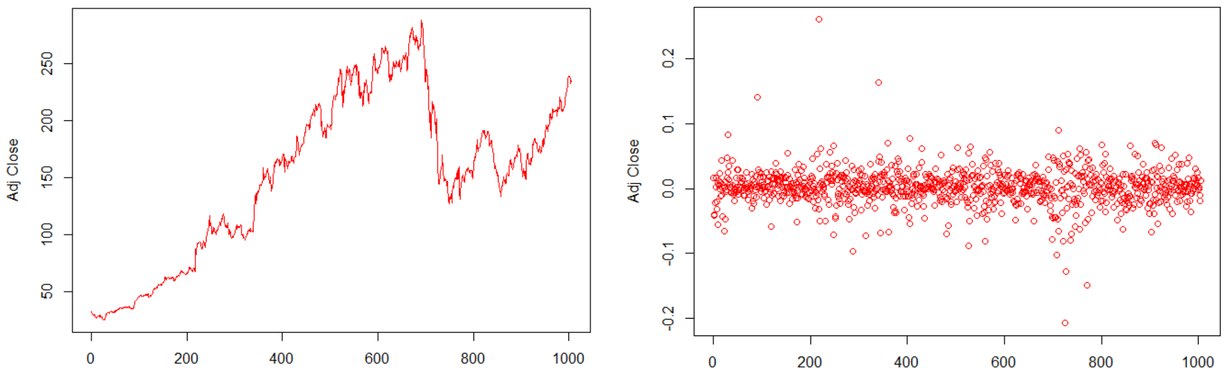

Figure 1. The NVDA Stock Adjusted Price and Log Return Distribution (2016.01~2019.12)

Time Series Model

Divide the data set, using the data from January 2, 2016 to December 31, 2018 as the training set with a total of **753**, and the data from January 1, 2018 to December 31, 2019 as the test set with a total of **252**.

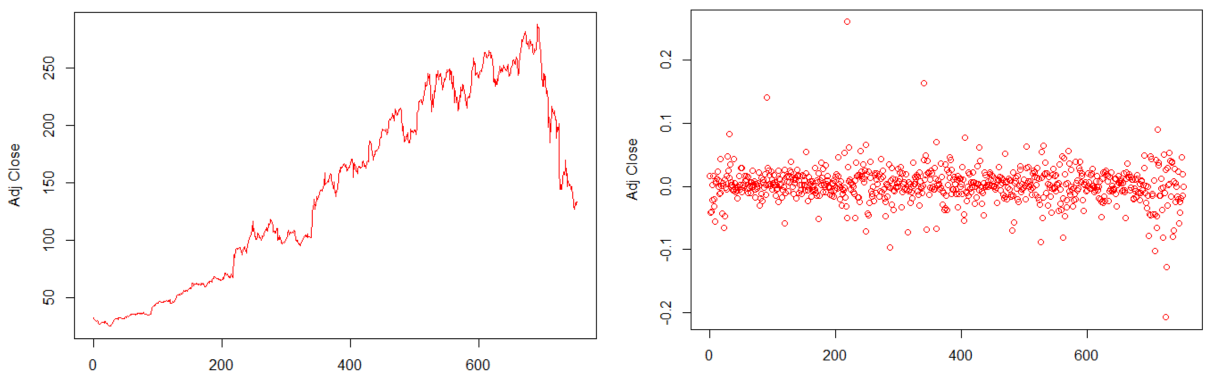

Figure 2. The NVDA Stock Adjusted Price and Log Return Distribution (2016.01~2018.12)

First use the ADF test to verify the stationarity of the sequence, the p value is **0.01** less than 0.05, so **reject** the null hypothesis that the sequence is **stationary**.

Then use the Ljung-Box test to check whether there is a lag correlation in the time series. Because the number of observations in the training set is 753, the lag parameter in the box test is ln(number of observations) approximately equal to 6. The p value in the box test is 0.1595 greater than 0.05, so **accept** the null hypothesis that the sequence is **independent**.

Through the above two tests, it can be determined that the log return series is a stationary time series, and there is no lag correlation, so no additional processing is required for the time series.

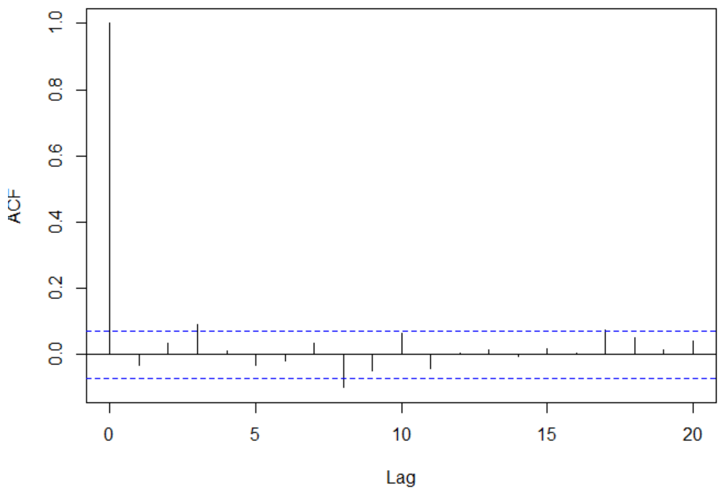

Figure 3. The ACF of Log Return

Choose the ARIMA model to fit the log return. Use the auto.arima(d,p,q) function to automatically search for the optimal model. Since the sequence is a stationary sequence, the d parameter equals to 1. From the output of auto. arima function, the best model is ARIMA (0,0,0), the details of output is as follows.

Table 2. The Output of ARIMA (0,0,0)

sigma^2 estimated as 0.0008154: log likelihood=1609.16							
AIC=-3216.32			AICc=-3216.32		BIC=-3211.7		
Training set error measures							
	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	0.002	0.029	0.019	100	100	0.673	-0.033

In general, if the model is suitable, the residuals of the model should meet the normal distribution with a mean of 0, and for any lag order, the residual autocorrelation coefficient should be zero. In other words, the residuals of the model should satisfy an independent normal distribution (that is, there is no correlation between the residuals).

Therefore, do normality test and independence test on the residual of ARIMA (0,0,0). From the Ljung-Box test output, the p value is 0.011, which means the residual sequence do not satisfy the independence assumption.

Because the residual does not satisfy the independence assumption, we specify the ARMA-GARCH model and fit with normal residuals.

$$\textbf{GARCH (1,1) Model:} \quad h_t = \omega + \alpha e^2_{\{t-1\}} + \beta h\_\{t-1\}$$

From the output, it is the estimated parameter of ARIMA(1,1) and GARCH(1,1) models were exhibited. It is seen that all parameters are significant. The log likelihood is 1651.893. In Weighted Ljung-Box Test on Standardized Residuals and Standardized Squared Residuals, the p value is too high to reject the null hypothesis that there is no serial correlation.

```
Conditional Variance Dynamics
-----------------------------------
GARCH Model      : sGARCH(1,1)
Mean Model       : ARFIMA(0,0,0)
Distribution     : norm

Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.004232    0.000903   4.6855    3e-06
omega   0.000213    0.000041   5.2217    0e+00
alpha1  0.405773    0.098742   4.1094    4e-05
beta1   0.443813    0.077320   5.7400    0e+00

Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.004232    0.001695   2.4967 0.012537
omega   0.000213    0.000105   2.0303 0.042321
alpha1  0.405773    0.344247   1.1787 0.238507
beta1   0.443813    0.224651   1.9756 0.048204

LogLikelihood : 1651.893
```

```
Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.2301  0.6315
Lag[2*(p+q)+(p+q)-1][2]    0.5627  0.6649
Lag[4*(p+q)+(p+q)-1][5]    2.7247  0.4596
d.o.f=0
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.2506  0.6167
Lag[2*(p+q)+(p+q)-1][5]    1.0155  0.8562
Lag[4*(p+q)+(p+q)-1][9]    1.4309  0.9610
d.o.f=2

Weighted ARCH LM Tests
------------------------------------
            Statistic Shape Scale P-Value
ARCH Lag[3]    0.2436 0.500 2.000  0.6216
ARCH Lag[5]    0.9200 1.440 1.667  0.7568
ARCH Lag[7]    1.0557 2.315 1.543  0.9043
```

Figure 4. The Output of GARCH Model with Normal

From the QQ-Plot for residuals, there are many observations have deviation from the theoretical model. For further confirmation, carry out the shapiro test to verify the normality. The p value is approximately equal to 0 and reject the normal hypothesis.

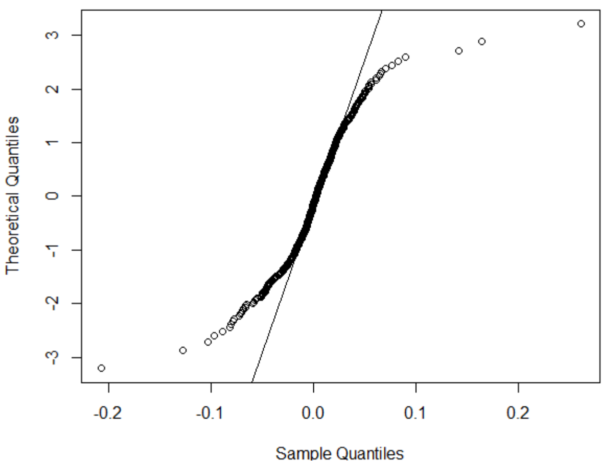

Figure 5. The Q-Q Plot with Normal Distribution

We refitted the ARMA-GARCH model with student t distribution. The model output is as follows. It is seen that all parameters are significant. In subsequent tests the p value is large enough, so the original hypothesis cannot be rejected.

```
Conditional Variance Dynamics
-----------------------------------
GARCH Model      : sGARCH(1,1)
Mean Model       : ARFIMA(0,0,0)
Distribution     : norm

Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.004232    0.000903   4.6855    3e-06
omega   0.000213    0.000041   5.2217    0e+00
alpha1  0.405773    0.098742   4.1094    4e-05
beta1   0.443813    0.077320   5.7400    0e+00

Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.004232    0.001695   2.4967 0.012537
omega   0.000213    0.000105   2.0303 0.042321
alpha1  0.405773    0.344247   1.1787 0.238507
beta1   0.443813    0.224651   1.9756 0.048204

LogLikelihood : 1651.893
```

```
Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.2301  0.6315
Lag[2*(p+q)+(p+q)-1][2]    0.5627  0.6649
Lag[4*(p+q)+(p+q)-1][5]    2.7247  0.4596
d.o.f=0
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.2506  0.6167
Lag[2*(p+q)+(p+q)-1][5]    1.0155  0.8562
Lag[4*(p+q)+(p+q)-1][9]    1.4309  0.9610
d.o.f=2

Weighted ARCH LM Tests
------------------------------------
            Statistic Shape Scale P-Value
ARCH Lag[3]    0.2436 0.500 2.000  0.6216
ARCH Lag[5]    0.9200 1.440 1.667  0.7568
ARCH Lag[7]    1.0557 2.315 1.543  0.9043
```

Figure 6. The Output of GARCH Model with std

Use the t distribution to fit the residuals, the degree freedom is 3.25. Then draw a Q-Q plot with df parameter is 3. From Figure7, it is clearly that majority of the observations are located in the line of y=x which shows good fit.

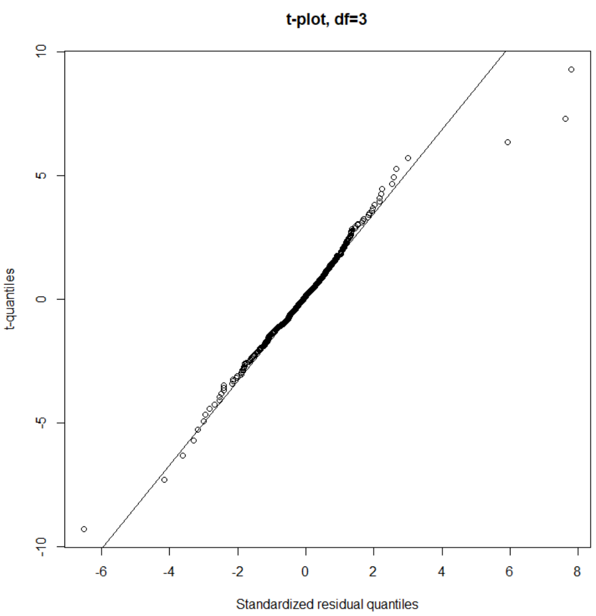

Figure 7. The Q-Q plot of t Distribution

Use ugarchroll function to predict the next 252 day’s log return. From Figure8, the prediction curve is very close to the actual curve, and the model prediction is quite effective.

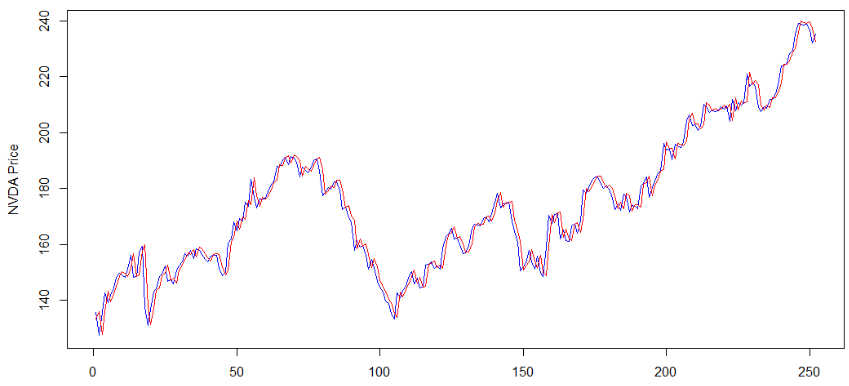

Figure 8. The Test Adjusted Price and Predict Adjusted Price Distribution

FNN, standard RNN, GRU and LSTM models

First, standardize the adjusted price.

$$x^* = \frac{x - \mu}{\sigma}$$

Next, divide the data set. Use the data after 2019 as the test set, and the data from 2016 to 2019 as the training set and validation set.

Finally, establish neural network model with different learning rate and choose the best one.

The learning rate is the hyperparameter that most affects the performance of the neural network model. It represents the speed of accumulation of information in the neural network over time. Compared with other hyperparameter learning rates, the effective capacity of the model is controlled in a more complicated way. When the learning rate is optimal, the effective capacity of the model is the largest. Therefore, choose three learning rate values of 0.1, 0.01 and 0.001 to train the model. The model output is as follows.

When the learning rate is 0.1, from Figure 9, there are different degrees of shock in both training loss and validation loss.

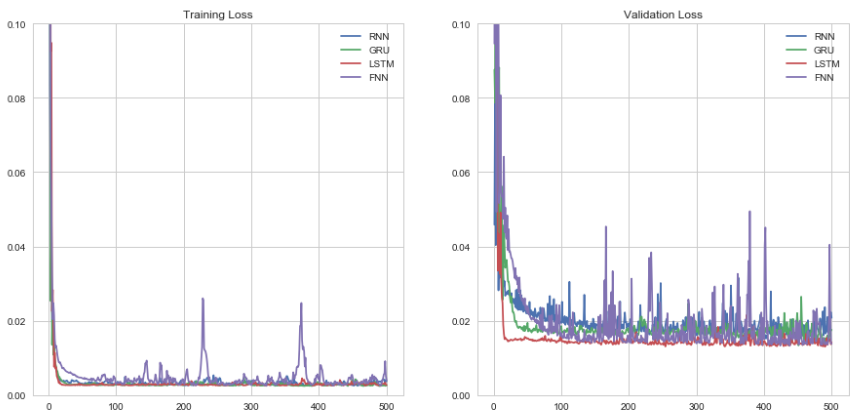

Figure 9. The Model Loss with 0.1 Learning Rate

Table 3. The Model MSE with 0.1 Learning Rate

Net	Train loss	Validation Loss	Test Loss
RNN	0.00384	0.02081	0.00400
GRU	0.00298	0.01758	0.00363
LSTM	0.00301	0.01416	0.00391
FNN	0.00531	0.01380	0.00434

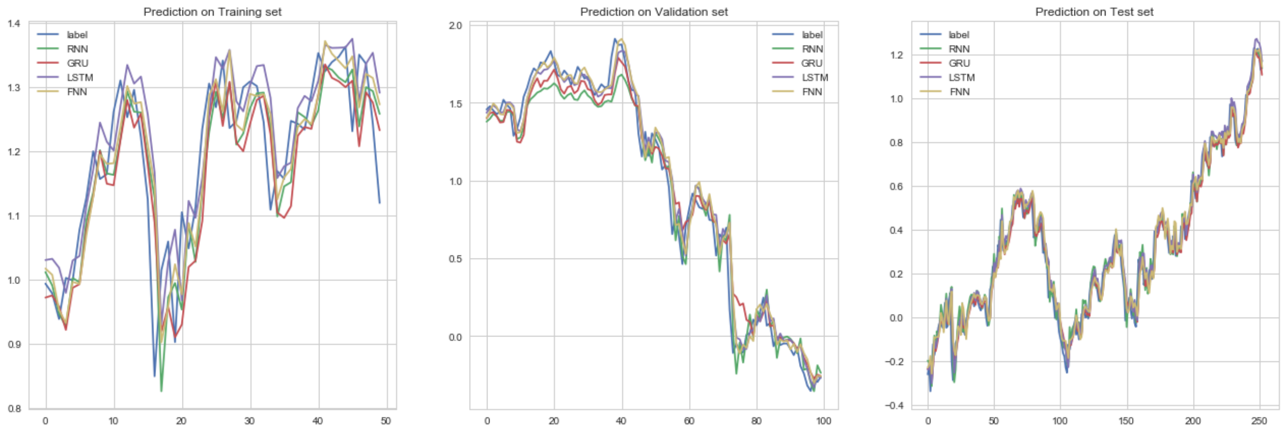

Figure 10. The Model Fitting Curve with 0.1 Learning Rate

When the learning rate is 0.001, from Figure 11, the loss curve is smooth, and finally reaches convergence when the number of iterations is around 250.

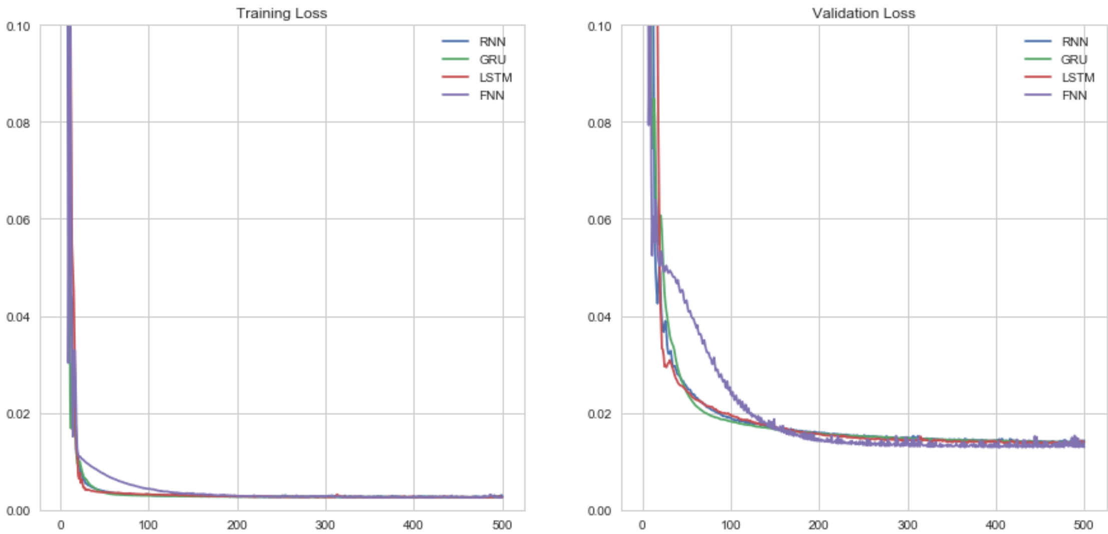

Figure 11. The Model Loss with 0.01 Learning Rate

Table 4. The Model MSE with 0.01 Learning Rate

Net	Train loss	Validation Loss	Test Loss
RNN	0.0028	0.0143	0.0038
GRU	0.0026	0.0140	0.0038
LSTM	0.0027	0.0139	0.0038
FNN	0.0027	0.0129	0.0039

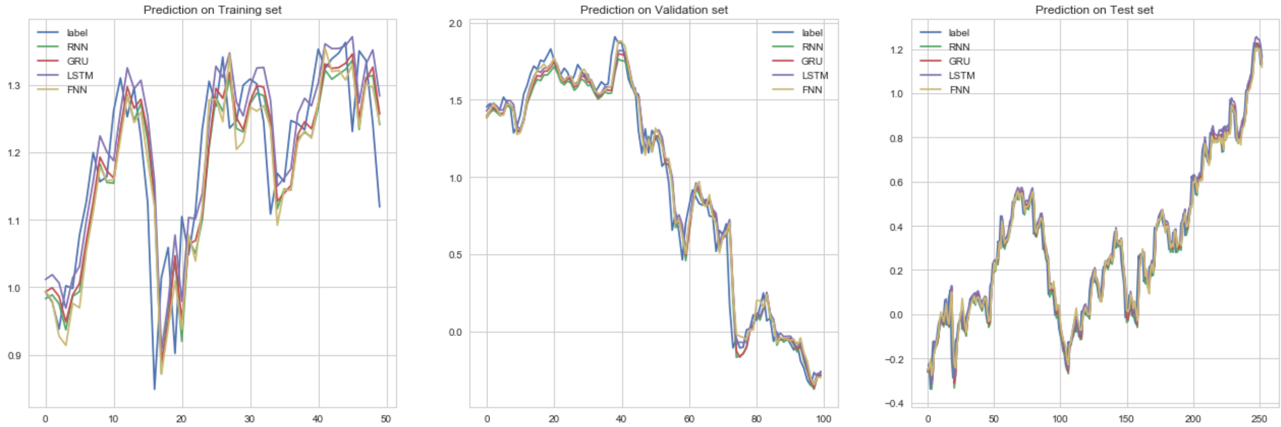

Figure 12. The Model Fitting Curve with 0.01 Learning Rate

When the learning rate is 0.001, from the Figure 13, it is clear that the training loss curve is smooth and finally converge. However, the validation loss does not converge.

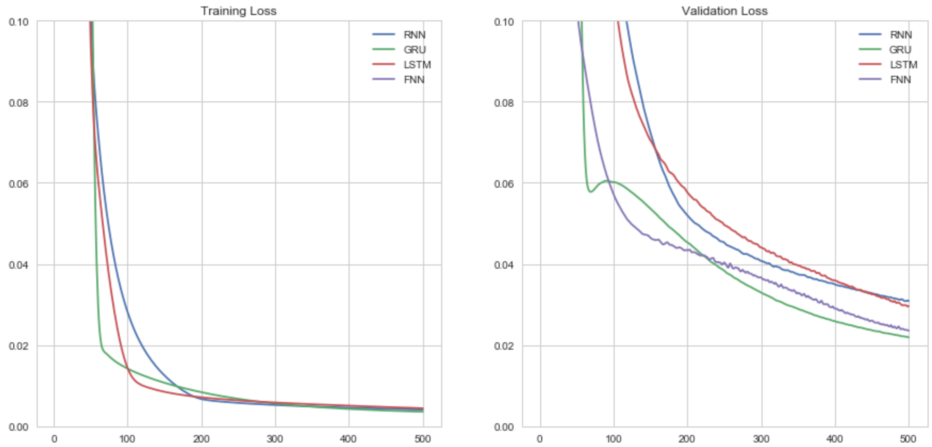

Figure 13. The Model Loss with 0.001 Learning Rate

Table 5. The Model MSE with 0.001 Learning Rate

Net	Train loss	Validation Loss	Test Loss
RNN	0.0041	0.0309	0.0072
GRU	0.0036	0.0219	0.0049
LSTM	0.0044	0.0296	0.0052
FNN	0.1238	0.0236	0.0062

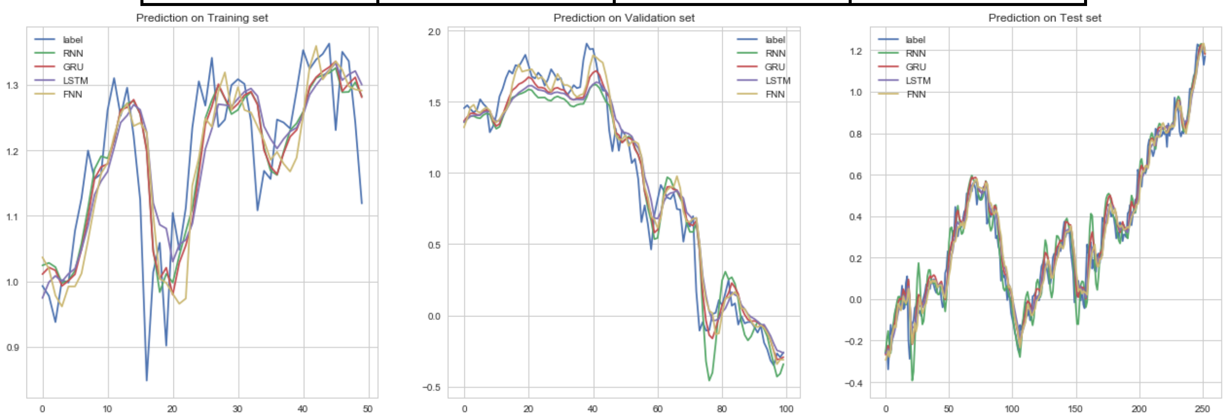

Figure 14. The Model Fitting Curve with 0.001 Learning Rate

By comprehensively comparing the three situations with different learning rates, we get two conclusions.

1. If the learning rate is small, the number of iterations required to reach convergence will be very high;

2. If the learning rate is large, each iteration may not reduce the result of the cost function, and may even exceed the local minimum and cause failure to converge.

Therefore, it was decided to set the learning rate to 0.01.

Prediction

Use mean squared error to evaluate two models.

$$MSE = \frac{SSE}{n} = \frac{1}{n}\sum_{i=1}^{m} w_i (y_i - \hat{y_i})^2$$

For the ARMA-GARCH model, the price of the test set is first standardized, and then predicted. The MSE is 0.00348.

For the neural network, all the models have the same level of loss on the test set. The best model is LSTM. The MSE of LSTM is 0.00375.

From the perspective of MSE indicators, there is no big gap in the fitting effect of the five models, and the ARMA-GARCH model is the smallest. From an explanatory point of view, the ARMA-GARCH model is more interpretable than the neural network model.

In summary, the ARMA-GARCH model is the best in predicting NVDA stock prices.

Table 6. The MSE of Models

Model	Test MSE
RNN	0.00380
GRU	0.00376
LSTM	0.00375
FNN	0.00385
ARMA-GARCH	0.00348