USING MACHINE LEARNING TO PREDICT THE EXPECTED RETURN OF ALIBABA

Team of
曲星凝，李钰，陳沛瑩，孔婷钰，黄雅丽

PART I
INTRODUCTION

INTRODUCTION OF CAPM

Capital asset pricing model (**CAPM**) is a model used to determine a theoretically appropriate required rate of return of an asset, to make decisions about adding assets to a well-diversified portfolio.

The **function** is as follow:

$$\bar{r_a} = r_f + \beta_a * (\bar{r_m} - \bar{r_f})$$

BASIC THOUGHT

ASSUME: the financial markets are effective

the investors are effectively dispersed as a whole

THEN the non systemic risk is not a concern.

R_f Constant

the model can be simplified as a **linear model** o$E(R_i)$ with respect t$E(R_m)$

THE POINT is using the historical data of market returns and stock returns to predictβ_i

.

DATA COLLECTING

Daily closing price of the Alibaba

Three major indexes of the United States: Standard Poole 500 stock index; Dow Jones industrial average; Nasdaq composite index

finance.yahoo.com

Logarithmic rate of return formula Rt=ln (Pt/Pt-1) the daily stock return.

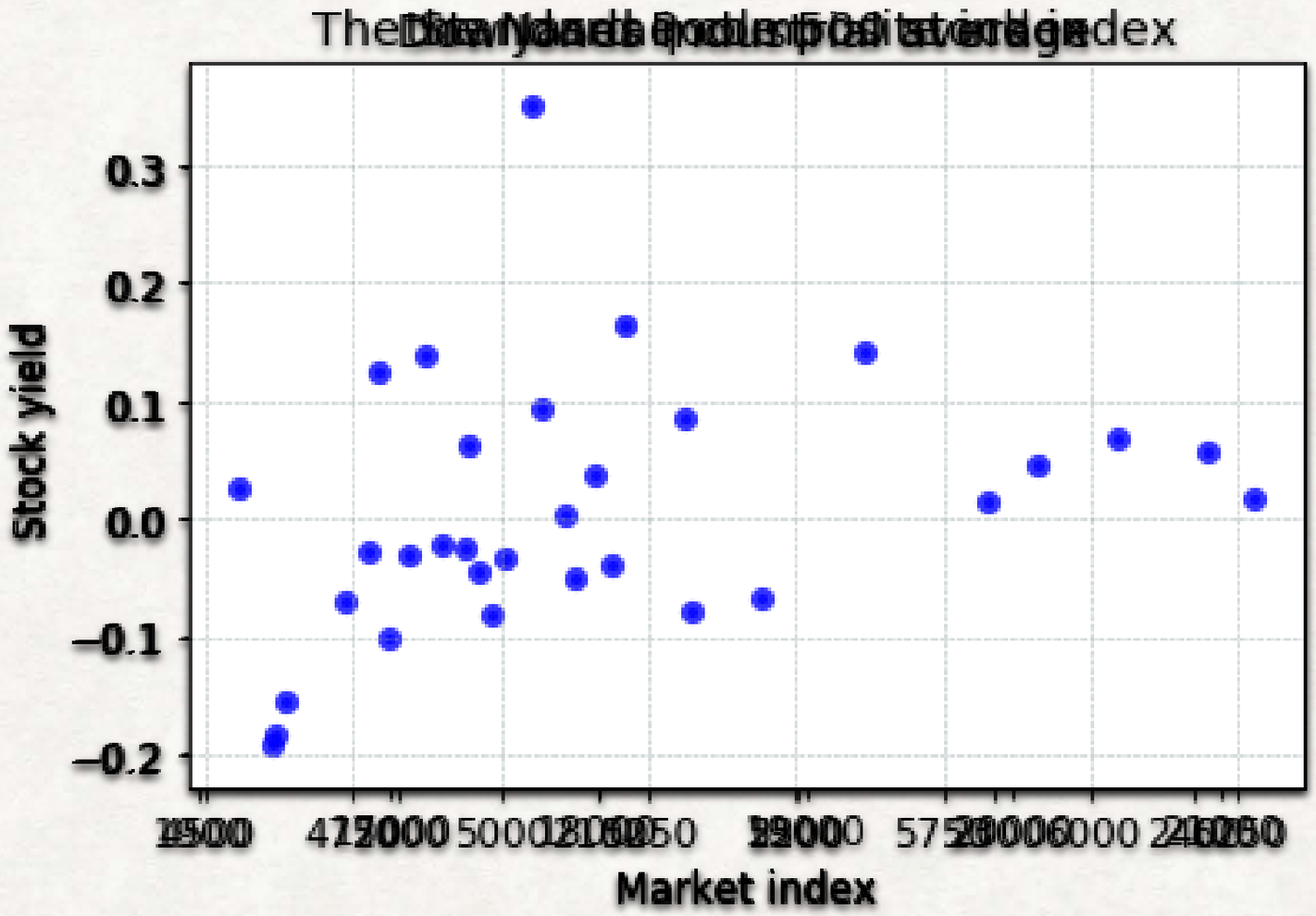
PART II
PROGRAMMING

PREPARATION

- Matplotlib

```
ra_and_rm=pd.DataFrame(pd.read_excel('ra_and_rm.xlsx'))
X = np.array(ra_and_rm[['rm']])
Y = np.array(ra_and_rm['ra'])
X.shape,Y.shape
plt.rc('font', family='STXihei', size=15)
plt.scatter(X,Y,60,color='blue',marker='o',linewidth=3,alpha=0.8)
plt.xlabel('Market index')
plt.ylabel('Stock yield')
plt.title('Market index and stock yield analysis')
plt.grid(color='#95a5a6',linestyle='--', linewidth=1,axis='both',alpha=0.4)
plt.show()
```

PREPARATION



TRAIN

- Ridge Regression and Lasso

```
X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.2,
random_state=0)
reg = linear_model.Ridge (alpha = 0.5)
reg.fit (X_train,y_train)
reg.coef_

X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.2,
random_state=0)
reg = linear_model.Lasso (alpha = 0.1)
reg.fit (X_train,y_train)
reg.coef_
```

TEST

```
reg.intercept_
reg.score(X_train,y_train)
y_predict=list(reg.predict(X_test))

Y_test=list(y_test)
p=((y_test - reg.predict(X_test)) **2).sum()
print(p)
print(y_predict)
print(Y_test)
```

PART III
RESULT ANALYZING

TEST RESULT

Ridge Regression	The s&p 500 index				Dow-Jones Average				Nasdaq Composite			
Predicting value	0.01426	0.00207	0.01471	0.00460	0.01000	0.00472	0.00990	0.00351	0.01890	0.00113	0.01975	0.00447
Actual value	0.00670	0.01210	0.00290	0.00150	0.00670	0.01210	0.00290	0.00150	0.00670	0.01210	0.0029	0.00150
Error variance	0.00031				0.00012				0.00056			

Lasso	The s&p 500 index				Dow-Jones Average				Nasdaq Composite			
Predicting value	0.00309	0.00309	0.00309	0.00309	0.00870	0.00442	0.00860	0.00343	0.01302	0.00186	0.01350	0.00395
Actual value	0.00670	0.01210	0.00290	0.00150	0.00670	0.01210	0.00290	0.00150	0.00670	0.01210	0.0029	0.00150
Error variance	0.0000968				0.0000995				0.00026			

ALPHA?

Ridge Regression	
α	Error variance
0.00001	0.0733922379615
0.01	0.0732631818682
0.1	0.0721750963988
1	0.0698033642465

Lasso	
α	Error variance
0.00001	0.0733923679755
0.01	0.0733923679704
1	0.0733923674649

PROBLEMS & CONCLUSION

- We assume that R_f is constant, so there can be a liner relationship between Market index and Stock yield
- To ignore the changes R_f we use short term datas, which can be a bit small.
- It’ s hard to get prediction of future market index.

THANK YOU FOR LISTENING

Special Thanks for our teacher BaoYang who has provided assistance and guide through our programming.