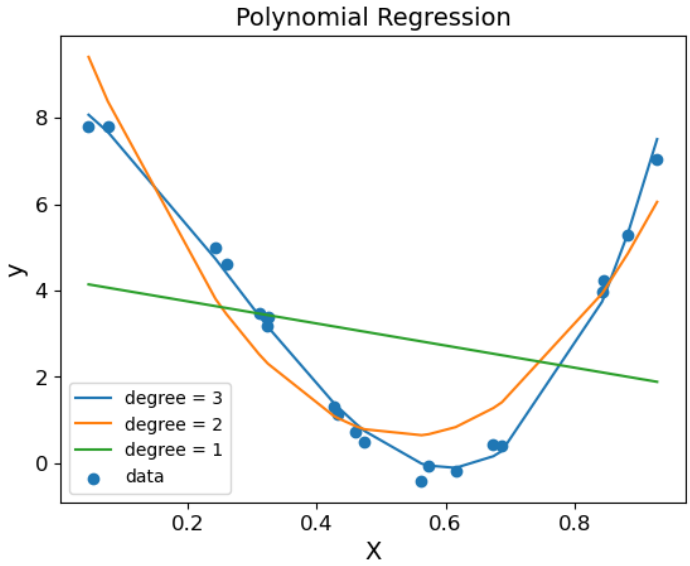
機器學習應用概論 HW2 r12631001 許喬淇

Problem 1

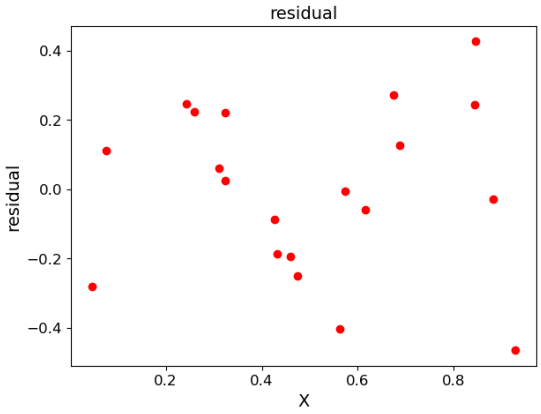
使用一至三階多項式函數進行擬和，三次多項式函數有較好之表現，其R^2約為0.992。

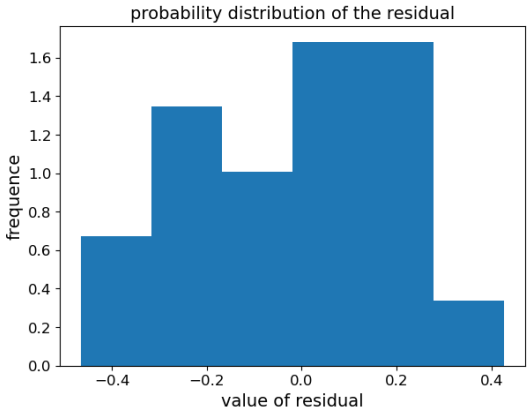


order	R^2
1st	0.0600
2nd	0.8969
3rd	**0.9920**

```
# 程式碼
regressor = make_pipeline(PolynomialFeatures(degree=n), LinearRegression())
regressor.fit(X, y)
plt.plot(X, regressor.predict(X),    label = "degree =”+str(n)+” 3")
score=regressor.score(X,y)
```

檢視 Residual 是否遵循最小平方法假設：Residual 與 X 獨立，且 Residual 之機率密度函數直方圖以零為中心(mean = 0)，並顯示鐘形曲線。(下圖以三階多項函數擬和)






```
# 程式碼
regressor = make_pipeline(PolynomialFeatures(degree=3), LinearRegression())
regressor.fit(X, y)
def residual(X,y):
    residual = []
    pred = regressor.predict(X)
    for i in range(0,20):
        residual.append(float(y[i]-pred[i]))
    return residual
```

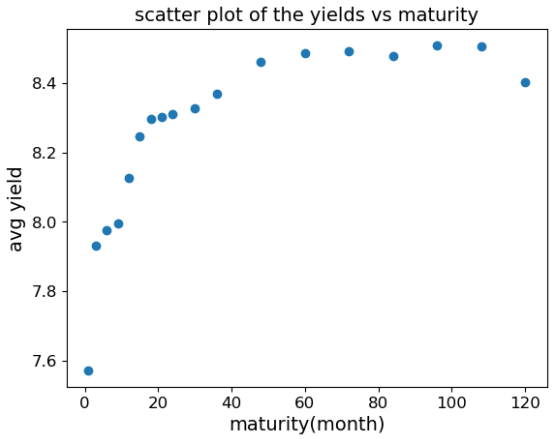
Problem 2

$\mathrm{Var}(\hat{w}) = \mathrm{Var}((X^TX)^{-1}X^T Y) = (X^TX)^{-1}X^T \mathrm{Var}(Y)\, X(X^TX)^{-1} = (X^TX)^{-1}\,\sigma^2$

Problem 3

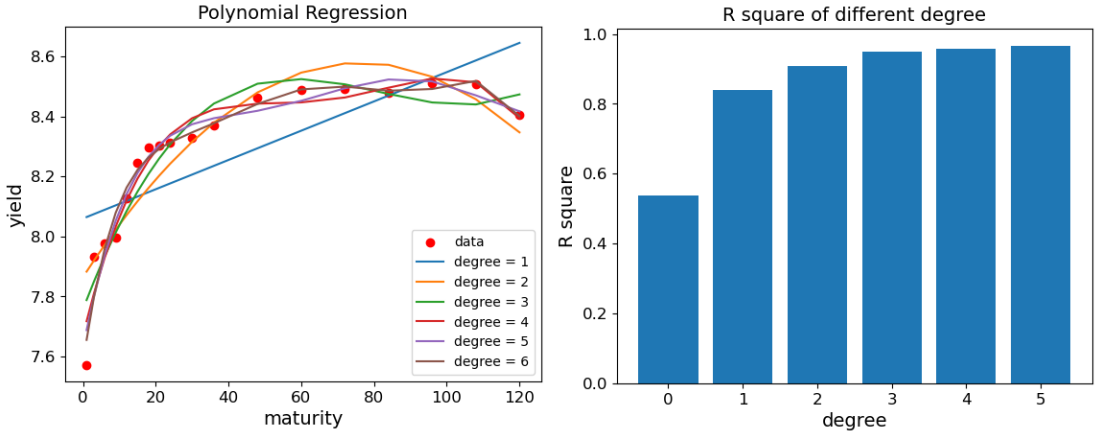
(a) Generate a scatter plot of the yields vs maturity.

收益隨著月數增加呈現上升的趨勢，在 100 個月時有最高的收益；但當到達 120 個月後呈現下降趨勢。



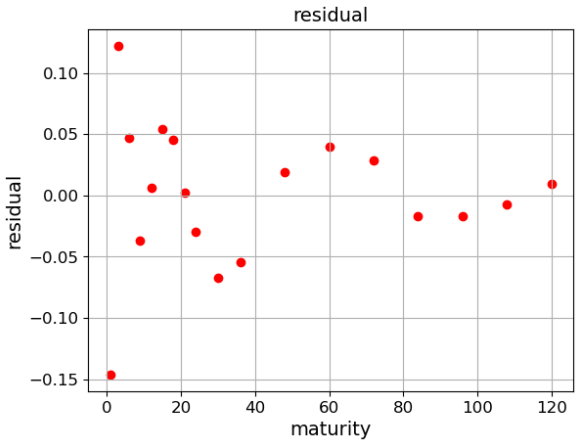
(b) You want to fit a polynomial regression model to the data. Since you don't know the order you need, you fit six polynomial models, with orders from 1 to 6. Plot the R^2 vs the polynomial order k.

隨著 degree 增加模型有越好的擬和結果，在 degree = 6 時 R^2 有最大值約為 0.9666。



(c) For the 4th-order polynomial model, draw a residual plot (vs maturity).

在 40 個月前之 residual 有震盪現象，40 個月後較趨緩且較接近 0。

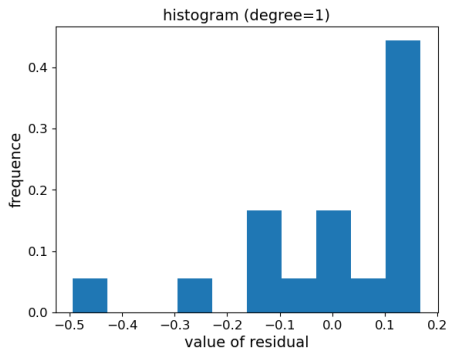


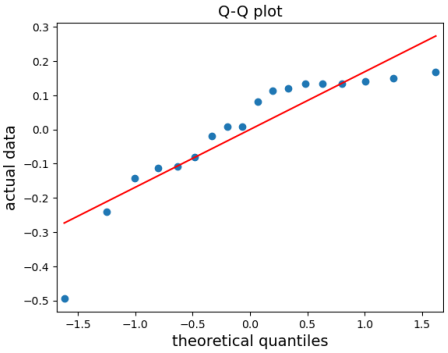
(d) Draw a histogram and a quantile-quantile (Q-Q) plot of the residuals.

Degree=4 時之機率分布直方圖較接近鐘形。

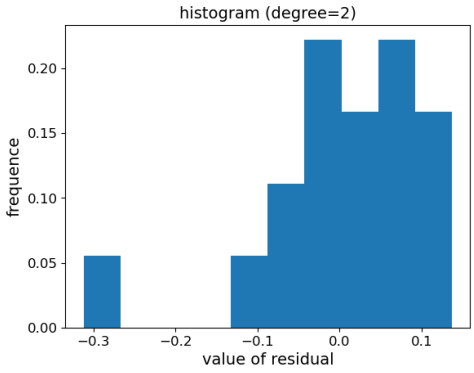
在 Q-Q 圖中，資料分布於 y = ax+b 為常態分布，分布於 y=x 則為標準常態分布，可以由下圖發現當 degree = 4 時，資料點分布情形最接近常態分佈。

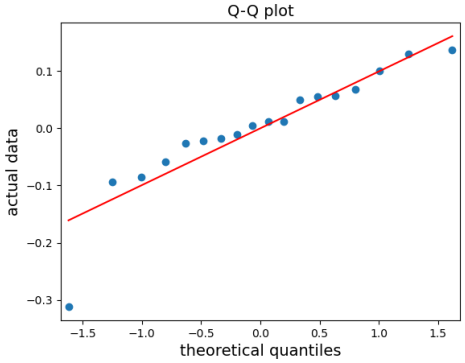
Degree = 1



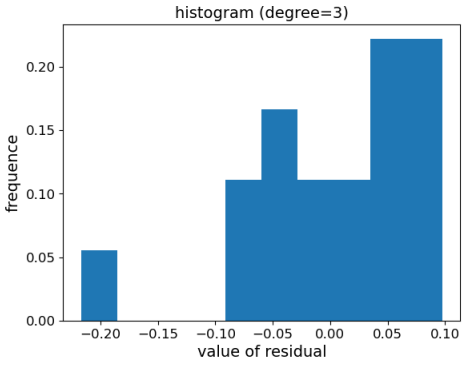


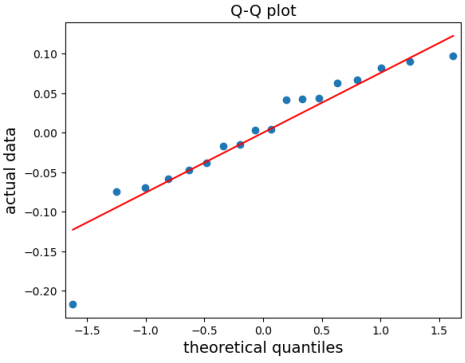
Degree = 2



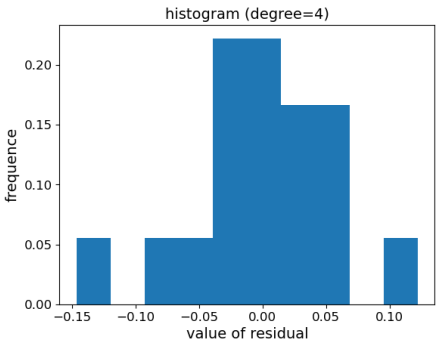


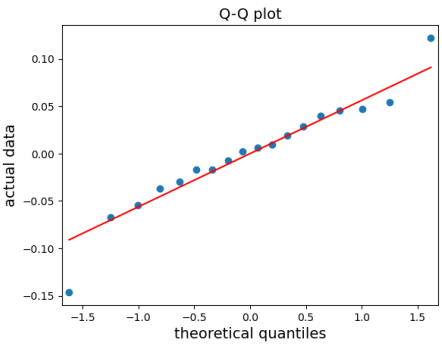
Degree = 3



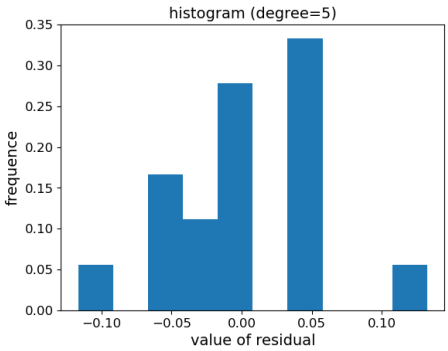


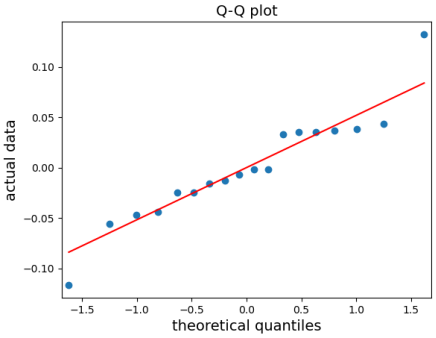
Degree = 4





Degree = 5





Degree = 6