Assignment #2－Regression

資工三 409410114 周述君

- Execution description: steps how to execute your codes.

➢ 使用 Colab，ctrl+F9 執行。

➢ 將參數存放於 Parameter 可以修改 num_points，num_points=15。

➢ 參數調整好後，Generate Samples 利用 linspace 函數調整 sample 產生的區間(np.linespace(0, 1)在 0 到 1 間均勻產生 num_points 個 sample , np.linespace(-3, 3)在-3 到 3 間均勻產生 num_points 個 sample)。

➢ 產生 sample 後，plot.scatter 將生成的 sample 繪製成二維散布圖。

➢ 調整 regression 中的 degree 顯示 linear regression(degree=1)和 polynomial regression(degree=5, degree=10, degree=14)的 fitting plot。

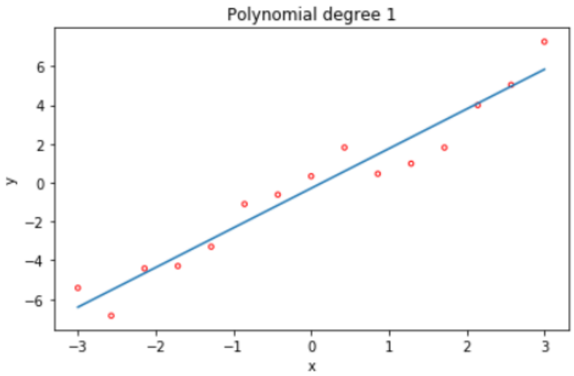
➢ 調整 lmda 來做正則化(lmda=0, lmda=0.001/m, lmda=1/m, lmda=1000/m)。

- Experimental results: As specified in the assignment

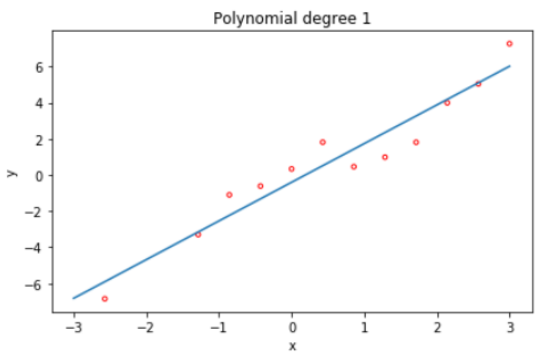
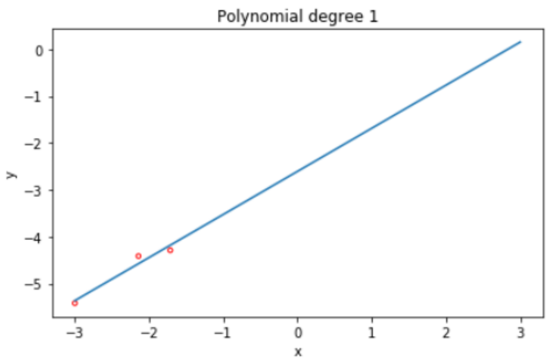
➢ Perform Linear Regression. Show the fitting plot, the training error, and the five-fold cross-validation errors.

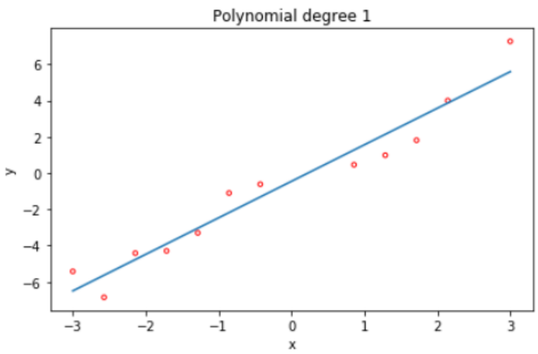
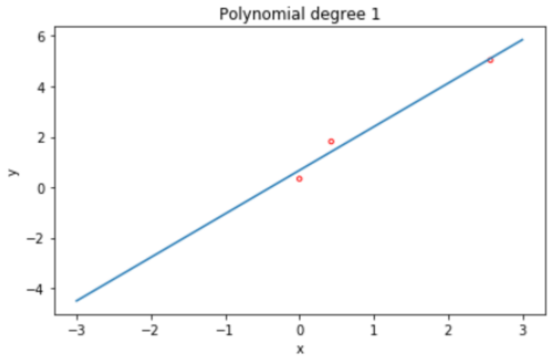
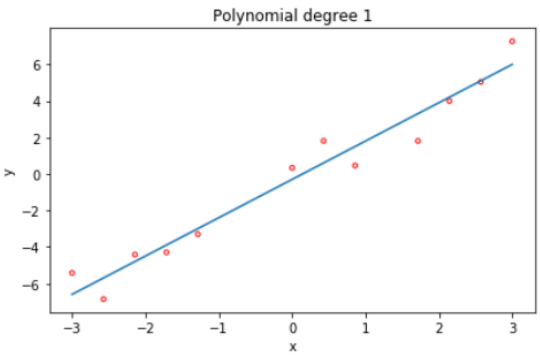
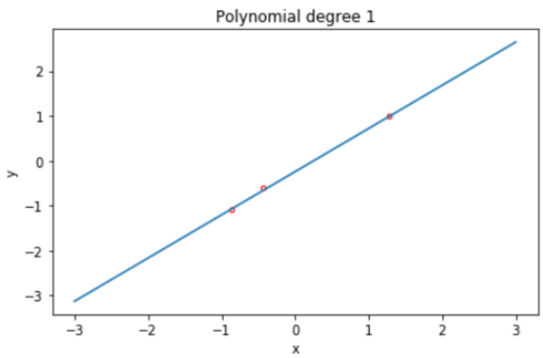
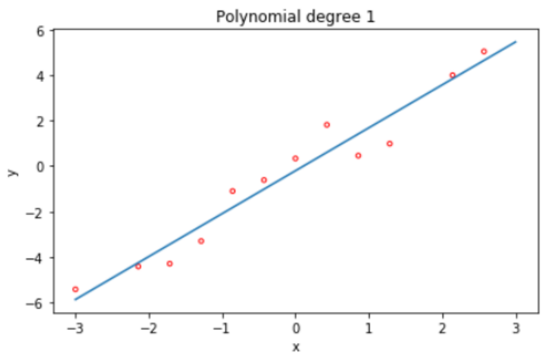
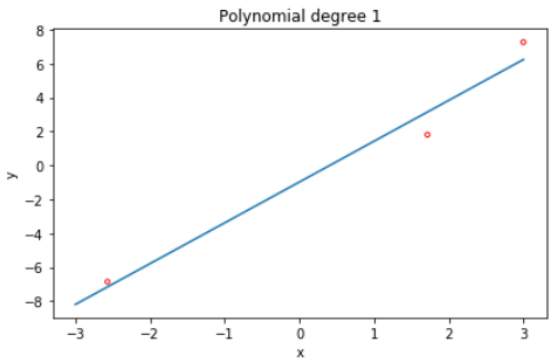
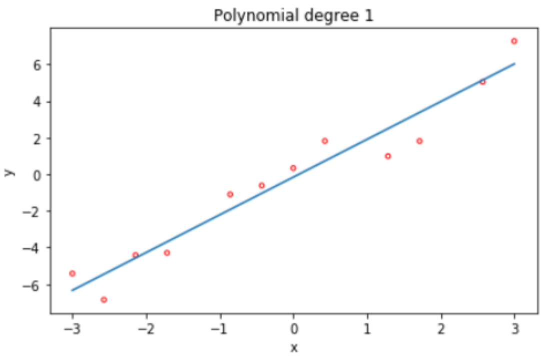
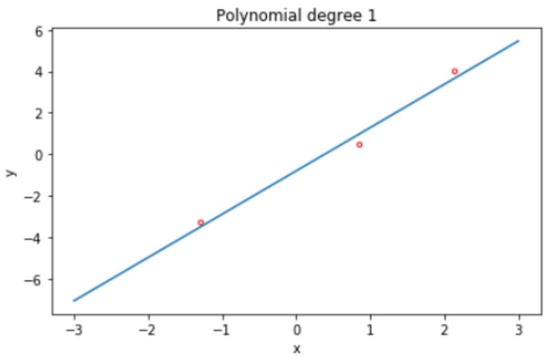
<u>Training error</u>

Degree=1, RMSE=0.6608466592234012



<u>Five-fold cross-validation errors</u>

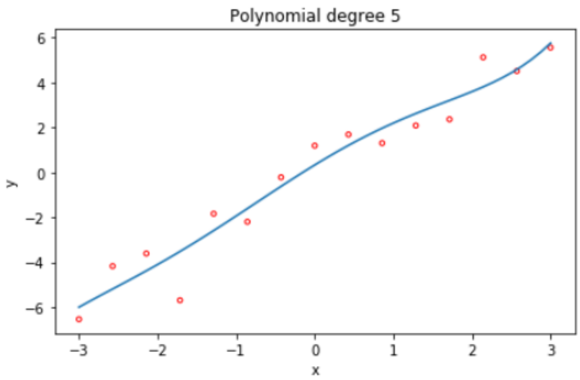
	training	valid
1	Degree=1, RMSE=0.6910131637989462 	Degree=1, RMSE=0.0826318079876958 

2	Degree=1, RMSE=0.6730182469722918  	Degree=1, RMSE=0.2159158078822646 
3	Degree=1, RMSE=0.6433321731453897  	Degree=1, RMSE=0.02136965063621975 
4	Degree=1, RMSE=0.5184623171237048  	Degree=1, RMSE=0.6996092850236884 
5	Degree=1, RMSE=0.6992285722498134  	Degree=1, RMSE=0.2646937947593224 

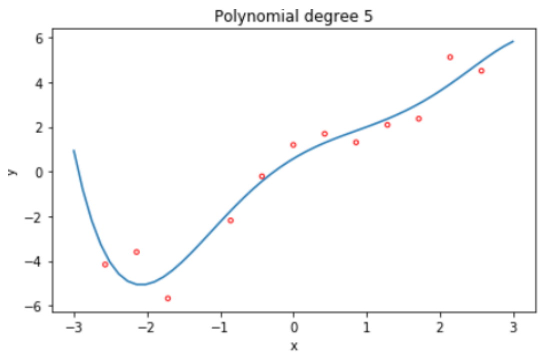
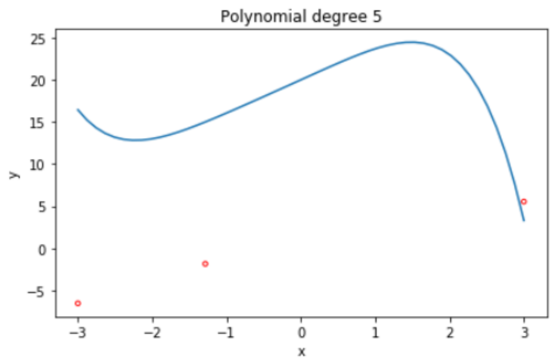
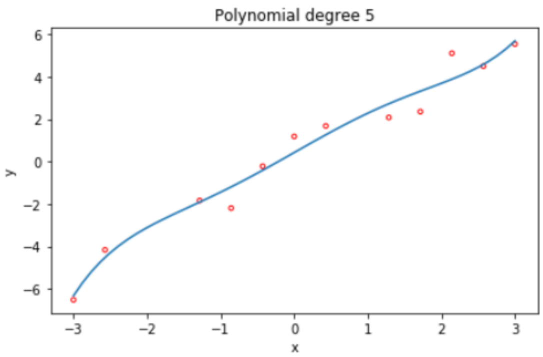
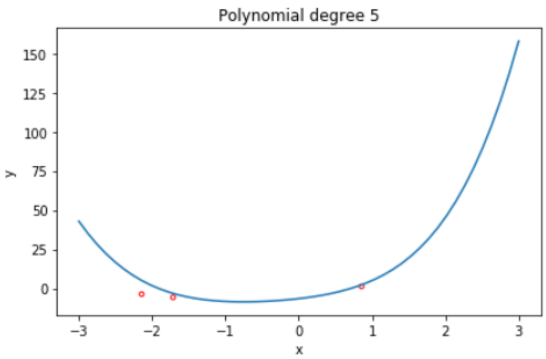
- Perform Polynomial Regression with degree 5, 10 and 14, respectively. For each case, show the fitting plot, the training error, and the five-fold cross-validation errors. (Hint: Arrange the polynomial regression equation as follows and solve the model parameter vector w.)
  - Degree=5

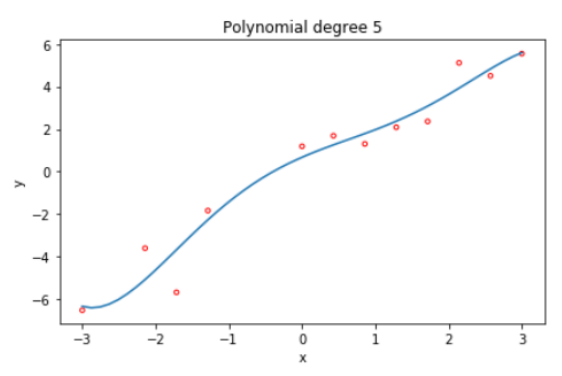
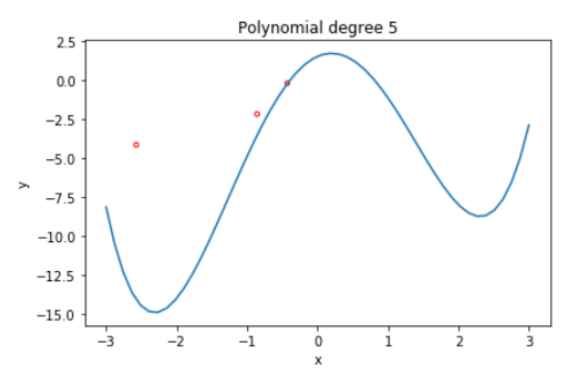
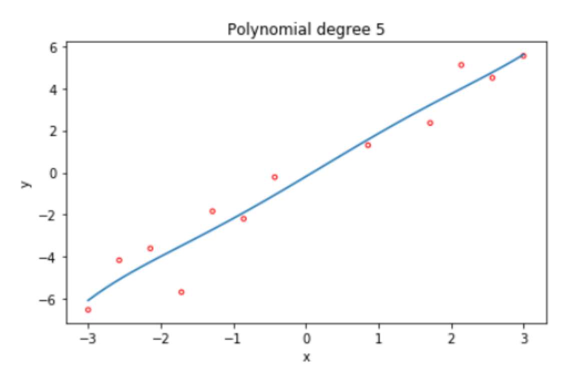
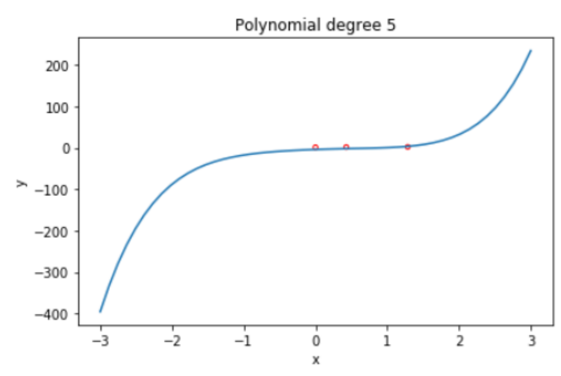
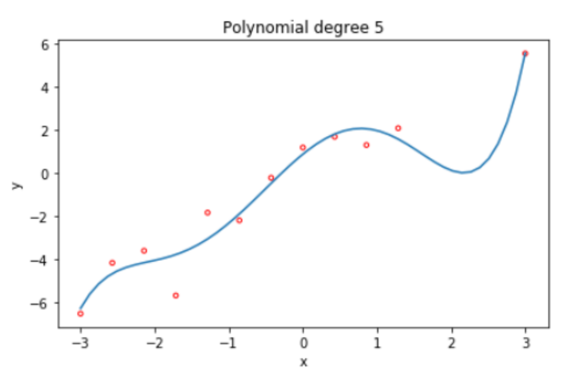
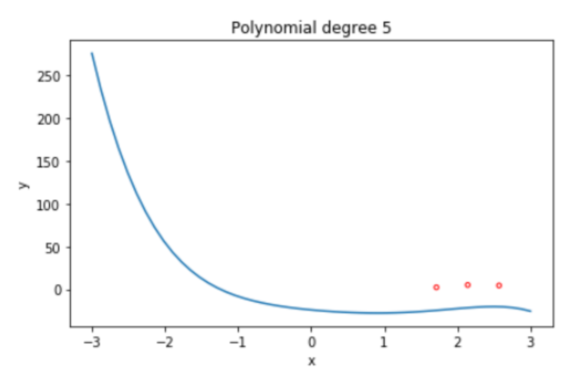
    <u>Training error</u>

    Degree=5, RMSE=0.6283404445398233



<u>Five-fold cross-validation errors</u>

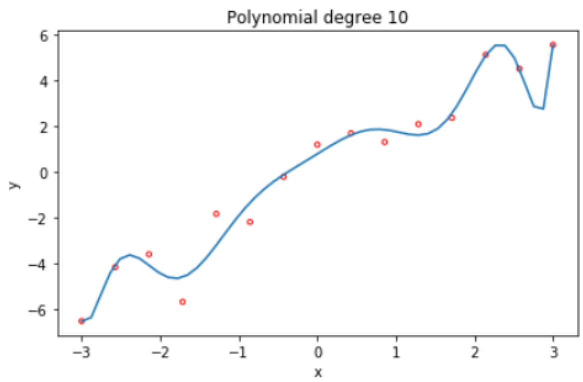
	training	**valid**
1	Degree=5, RMSE=0.5298677053453411	Degree=5, RMSE=11.660247515332363
		
2	Degree=5, RMSE=0.4491836894065976	Degree=5, RMSE=3.718096464816144
		
3	Degree=5, RMSE=0.6199437903881957	Degree=5, RMSE=4.086817144242787

		
4	Degree=5, RMSE=0.6458423938667138	Degree=5, RMSE=2.742977409302637
		
5	Degree=5, RMSE=0.5327445180295234	Degree=5, RMSE=18.476768791115596
		

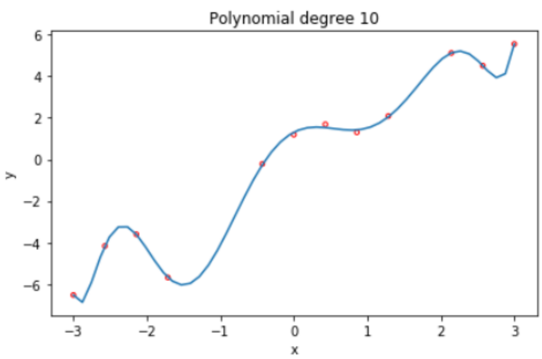
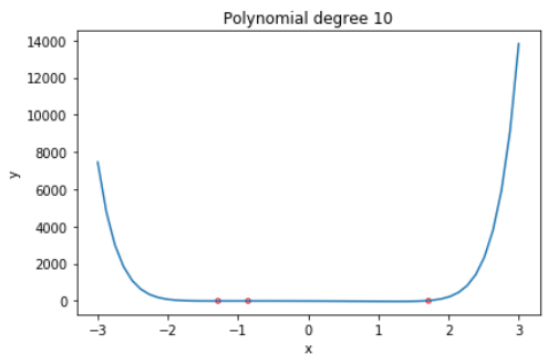
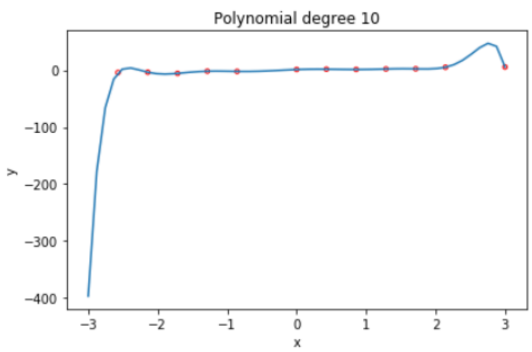
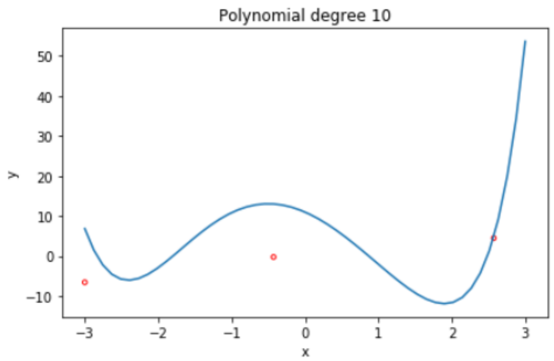
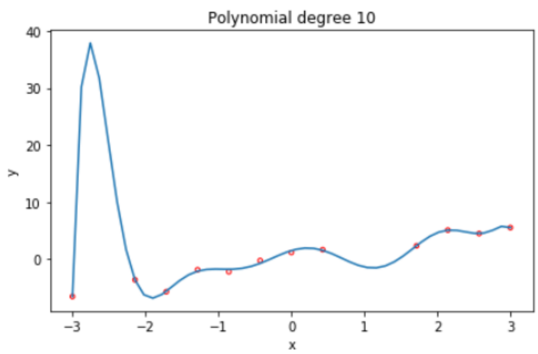
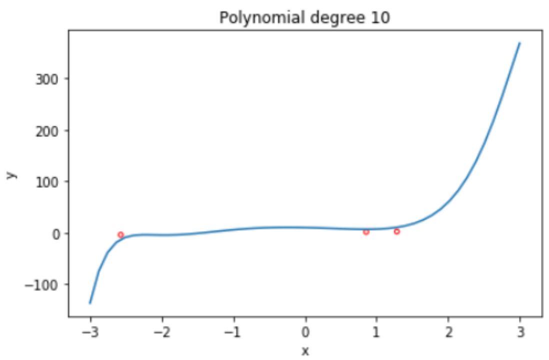
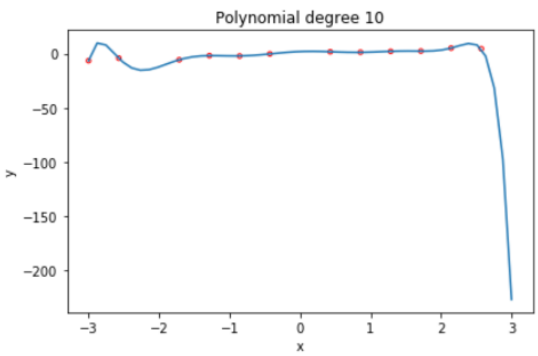
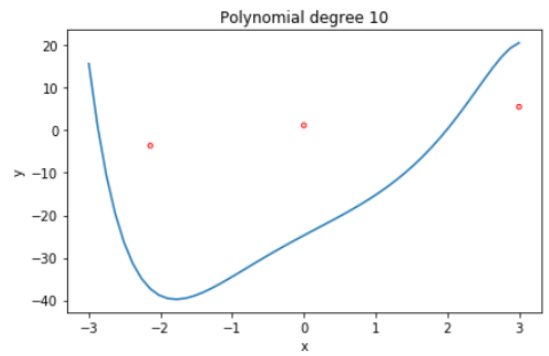
- Degree=10

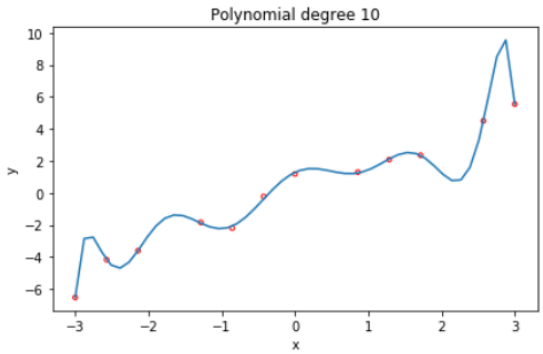
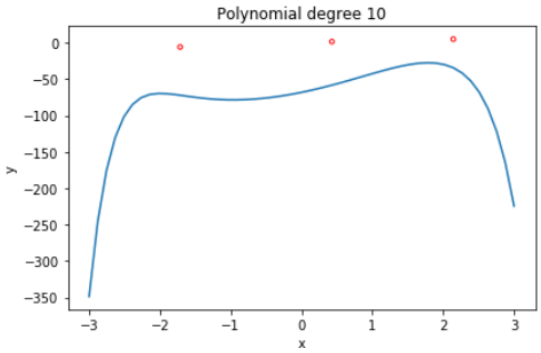
  <u>Training error</u>

  Degree=10, RMSE=0.3975256867402641



Five-fold cross-validation errors

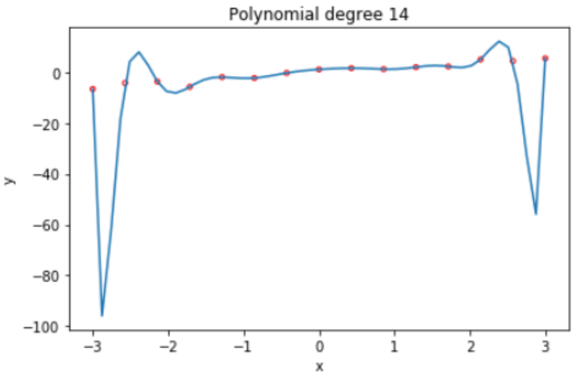
	training	**valid**
1	Degree=10, RMSE=0.048156581435724	Degree=10, RMSE=5.381509451890489
		
2	Degree=10, RMSE=0.0328826889355411	Degree=10, RMSE=7.689798264979682
		
3	Degree=10, RMSE=0.1636742848310417	Degree=10, RMSE=5.395815093751934
		
4	Degree=10, RMSE=0.0347989531191255	Degree=10, RMSE=18.409119755962415
		

5	Degree=10, RMSE=0.0498893169748178	Degree=10, RMSE=39.83707536208904
		

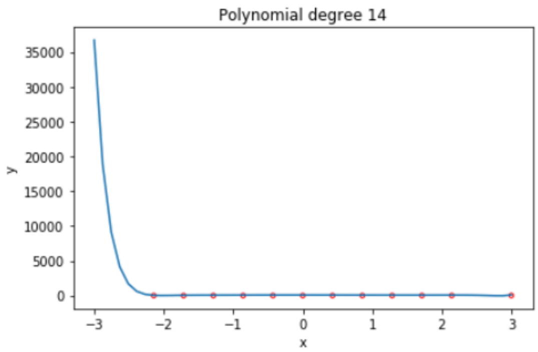
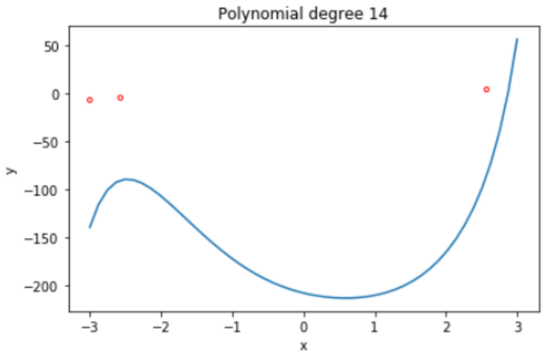
- Degree=14

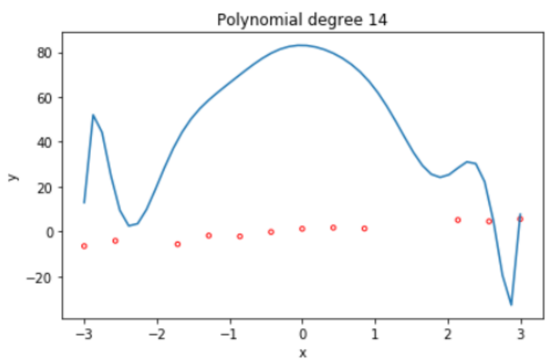
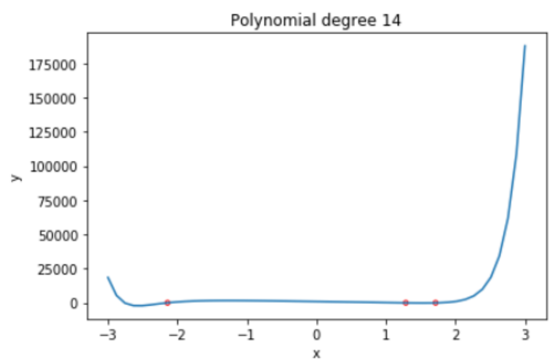
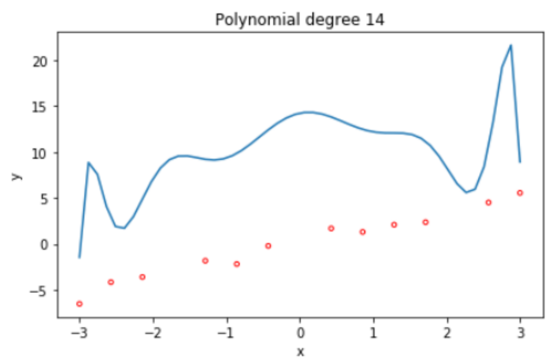
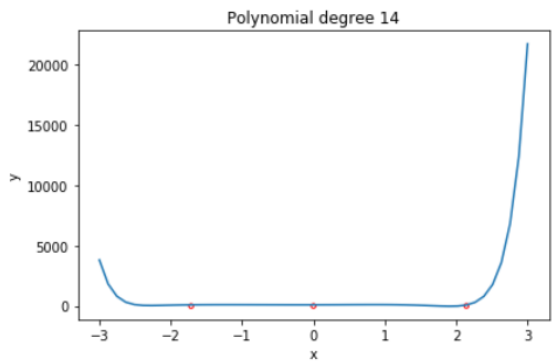
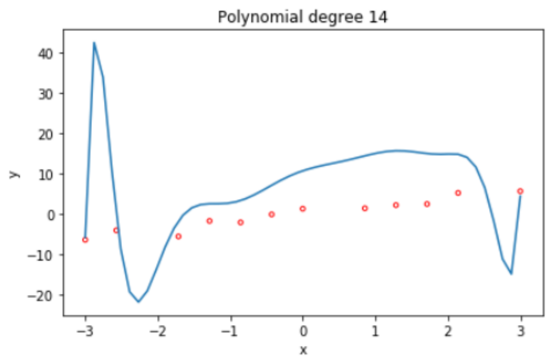
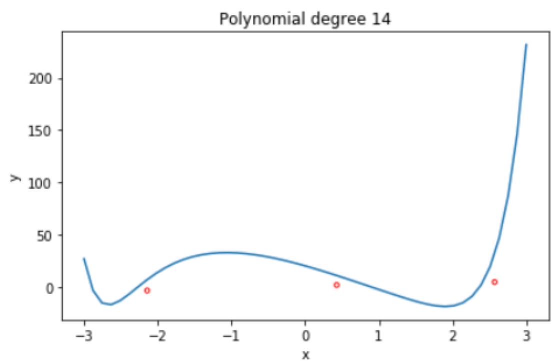
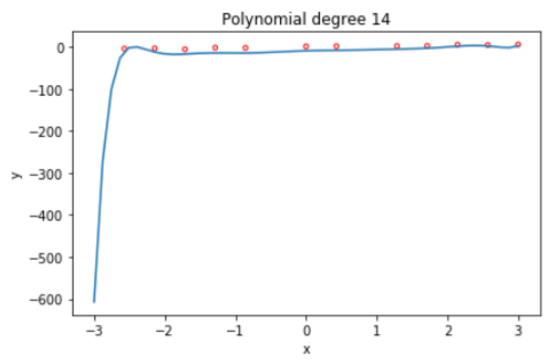
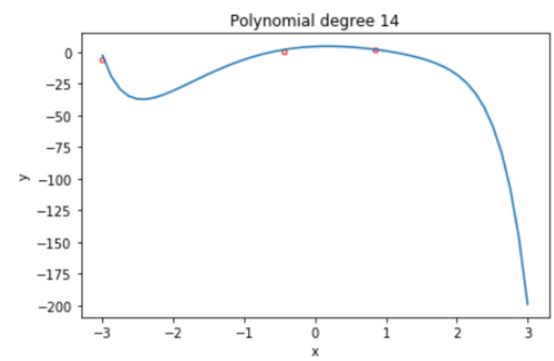
  Training error

  Degree=14, RMSE=7.136053752255e-06



Five-fold cross-validation errors

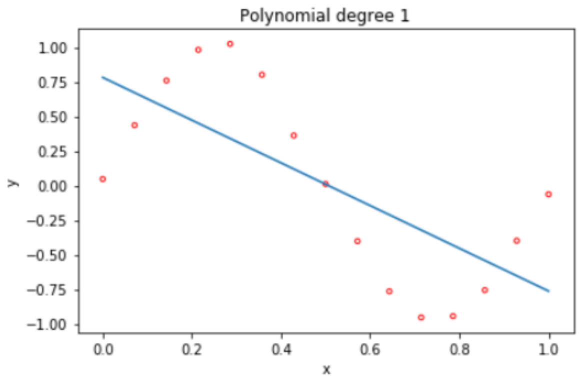
	training	**valid**
1	Degree=14, RMSE=45.23296143162496	Degree=14, RMSE=74.34497108158169
		
2	Degree=14, RMSE=38.82962743875646	Degree=14, RMSE=24.01688485555414

		
3	Degree=14, RMSE=6.673386260196252	Degree=14, RMSE=66.16691285242946
		
4	Degree=14, RMSE=5.8527239630012495	Degree=14, RMSE=12.35147984701117
		
5	Degree=14, RMSE=6.048017752680724	Degree=14, RMSE=1.807152057472771
		

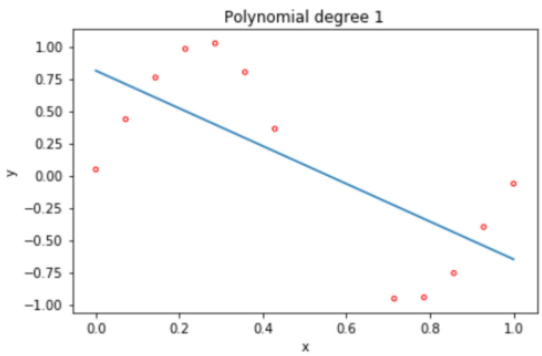
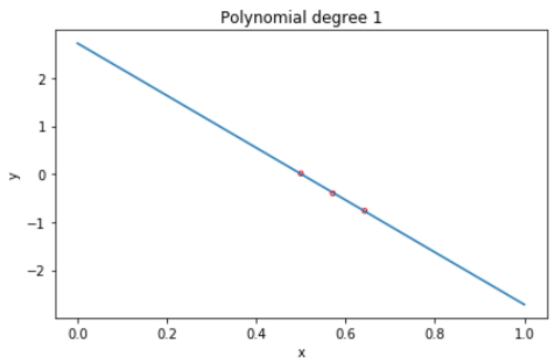
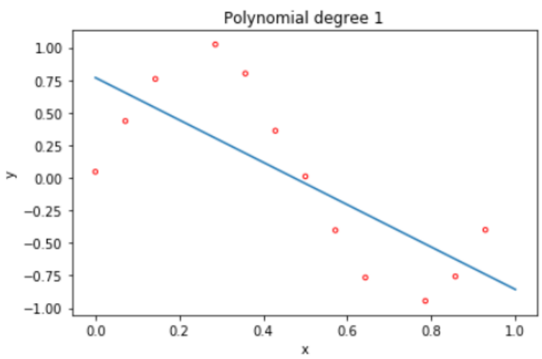
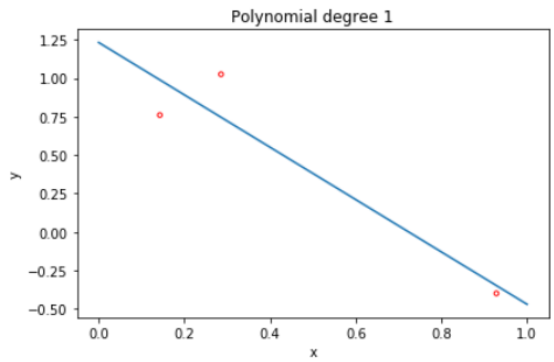
- Change the model to y = sin(2πx) +εwith the noise ε~ N(0, 0.04) and (equal spacing) x ∈ [0, 1]. Then repeat those stated in 2) and 3). Compare the results with linear/polynomial regression on different datasets.
  - Degree=1

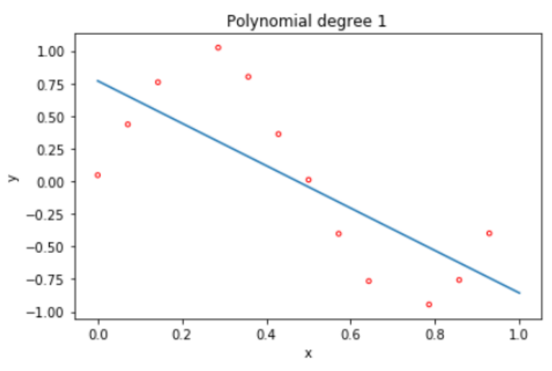
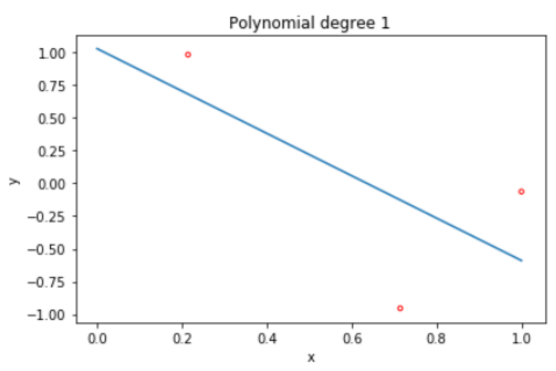
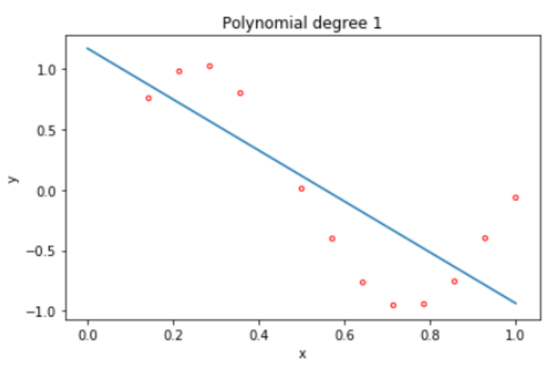
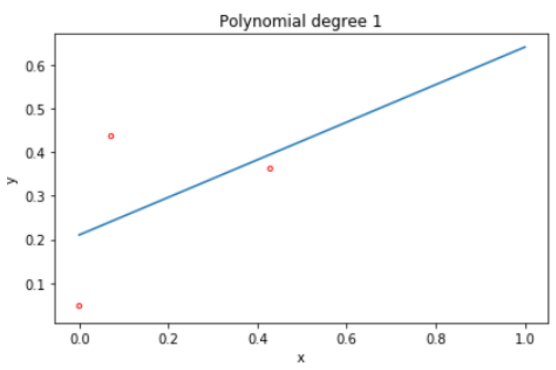
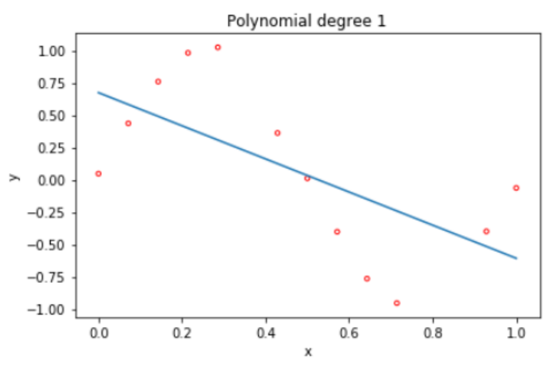
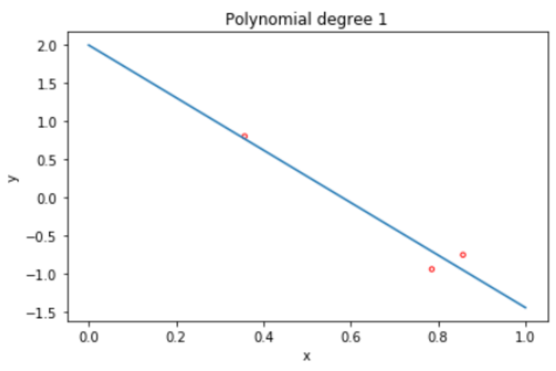
    <u>Training error</u>

    Degree=1, RMSE=0.3381117366078396



<u>Five-fold cross-validation errors</u>

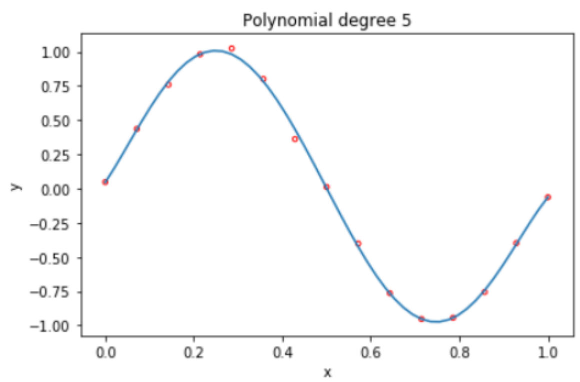
<table>
<tr><th></th><th>training</th><th>valid</th></tr>
<tr><td>1</td><td>Degree=1, RMSE=0.3513241594046753</td><td>Degree=1, RMSE=0.00841157733406936</td></tr>
<tr><td></td><td>
</td><td>
</td></tr>
<tr><td>2</td><td>Degree=1, RMSE=0.33763457280060355</td><td>Degree=1, RMSE=0.1488449843647542</td></tr>
<tr><td></td><td>
</td><td>
</td></tr>
<tr><td>3</td><td>Degree=1, RMSE=0.30421926826372075</td><td>Degree=1, RMSE=0.4184455303571044</td></tr>
</table>

<table>
<tr><td></td><td>
</td><td>
</td></tr>
<tr><td>4</td><td>Degree=1, RMSE=0.319042404406932</td><td>Degree=1, RMSE=0.10478813807483199</td></tr>
<tr><td></td><td>
</td><td>
</td></tr>
<tr><td>5</td><td>Degree=1, RMSE=0.33607727847437785</td><td>Degree=1, RMSE=0.12657315863893567</td></tr>
<tr><td></td><td>
</td><td>
</td></tr>
</table>

- Degree=5

<u>Training error</u>

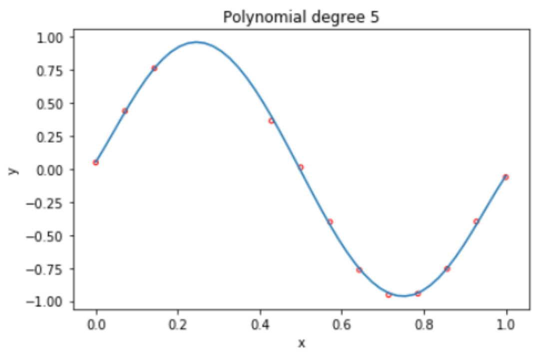
```
Degree=5, RMSE=0.017025404734936914
```



<u>Five-fold cross-validation errors</u>

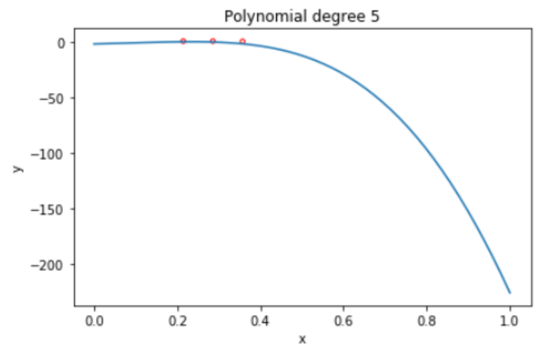
<table>
<tr><th></th><th>training</th><th>valid</th></tr>
<tr><td>1</td><td>

`Degree=5, RMSE=0.011951427621339135`



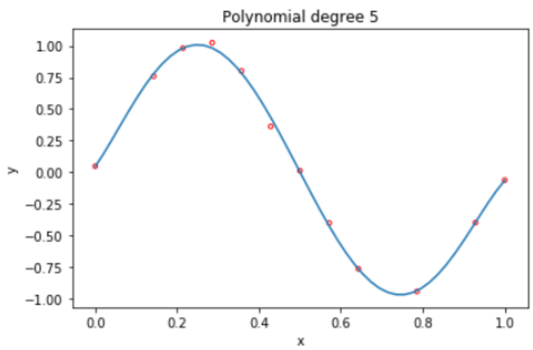
</td><td>

`Degree=5, RMSE=0.9824565288713851`



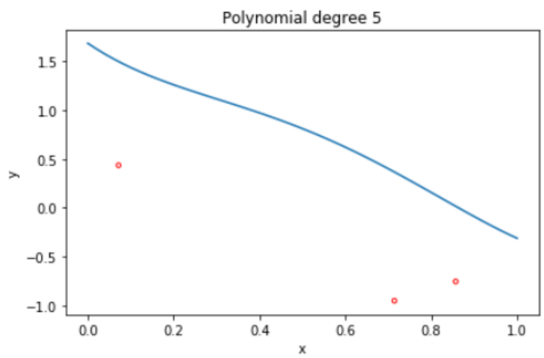
</td></tr>
<tr><td>2</td><td>

`Degree=5, RMSE=0.018657195409370313`



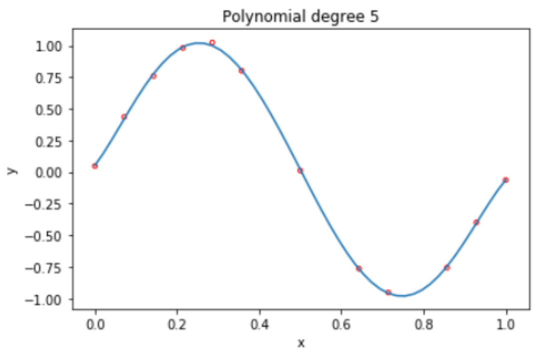
</td><td>

`Degree=5, RMSE=0.7608216066334113`



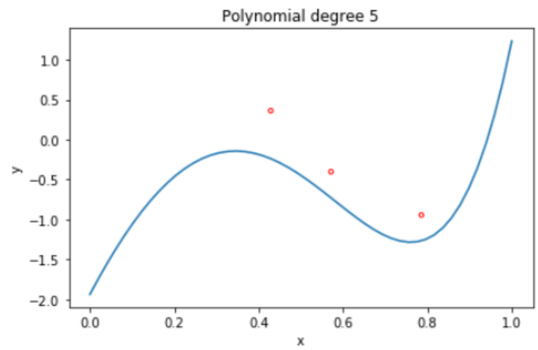
</td></tr>
<tr><td>3</td><td>

`Degree=5, RMSE=0.0079938152690887`



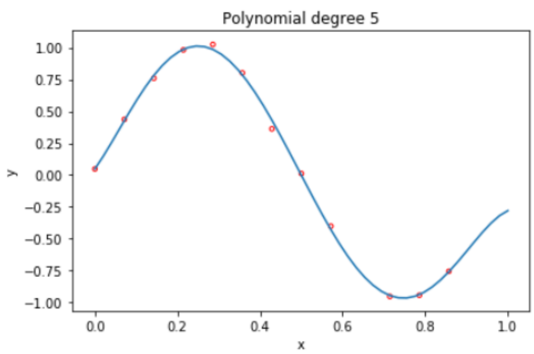
</td><td>

`Degree=5, RMSE=0.30861245183275676`



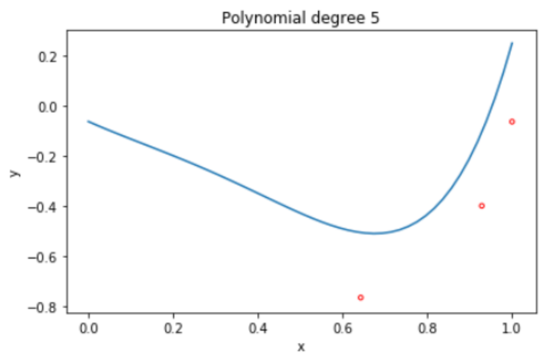
</td></tr>
<tr><td>4</td><td>

`Degree=5, RMSE=0.018468231922172892`

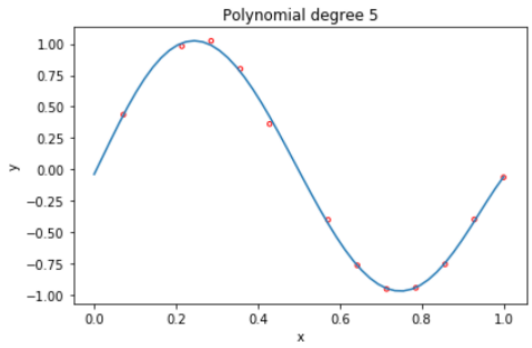
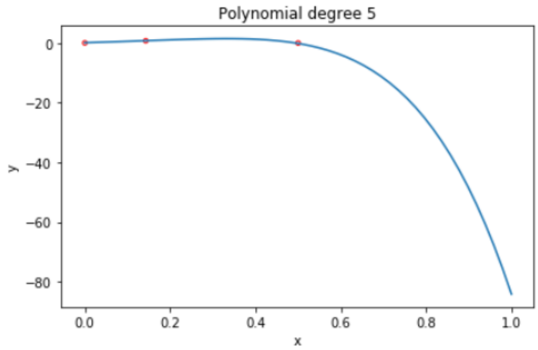


</td><td>

`Degree=5, RMSE=0.20483656554802698`



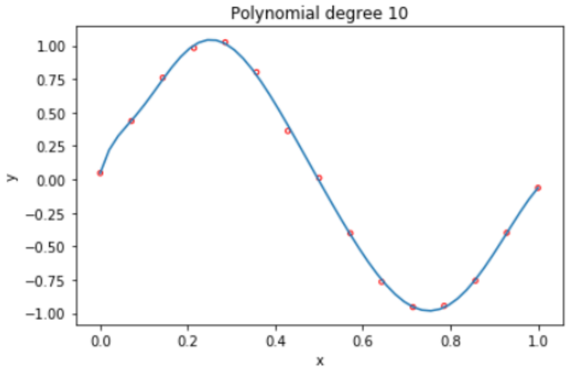
</td></tr>
</table>

5	Degree=5, RMSE=0.01716844516456322  	Degree=5, RMSE=0.06578992524029013 

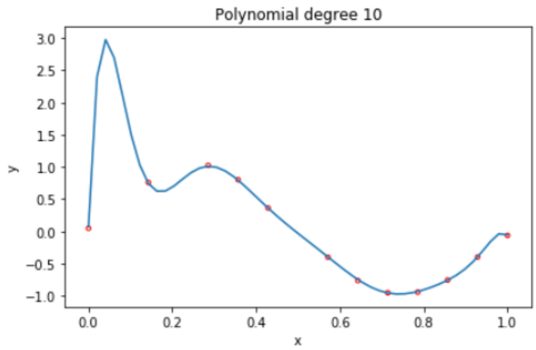
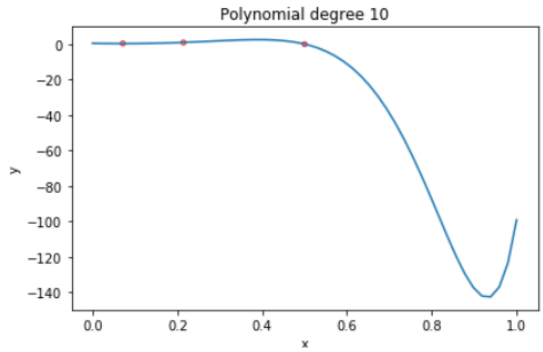
- Degree=10

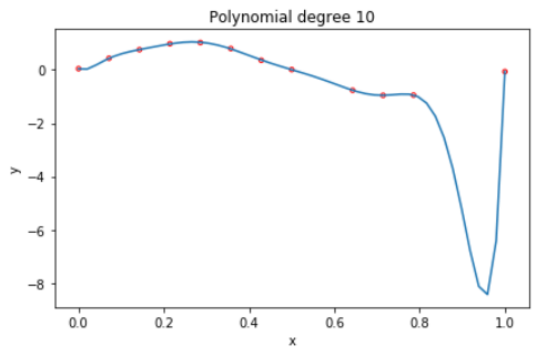
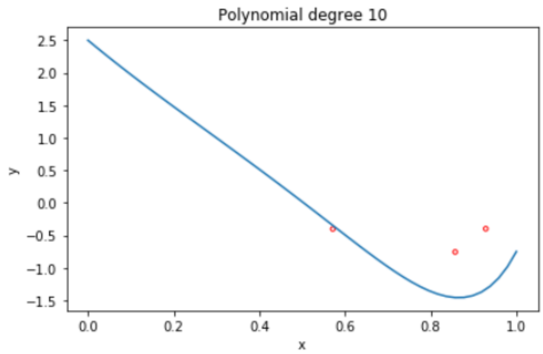
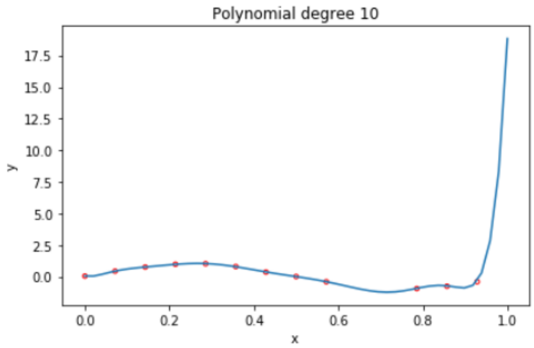
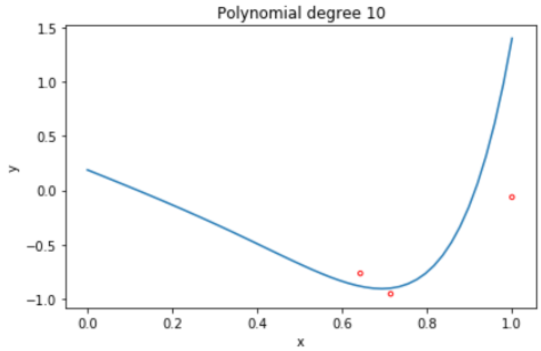
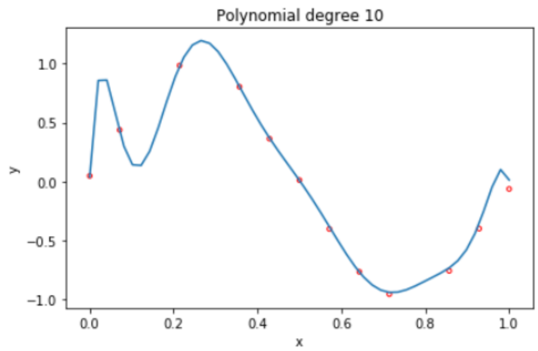
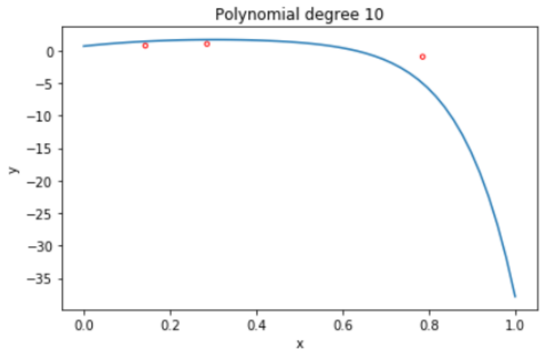
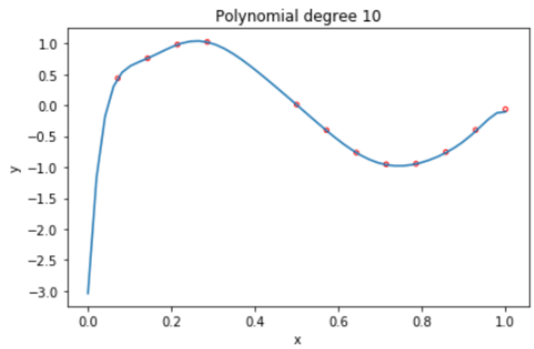
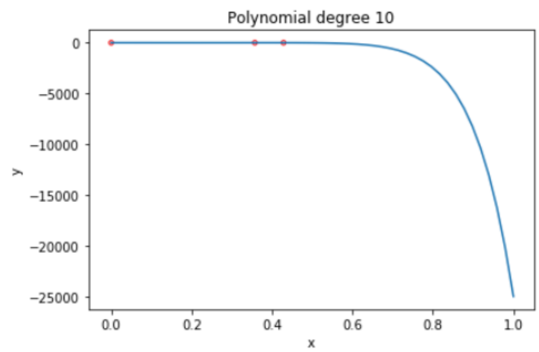
<u>Training error</u>

Degree=10, RMSE=0.011747387744402643



<u>Five-fold cross-validation errors</u>

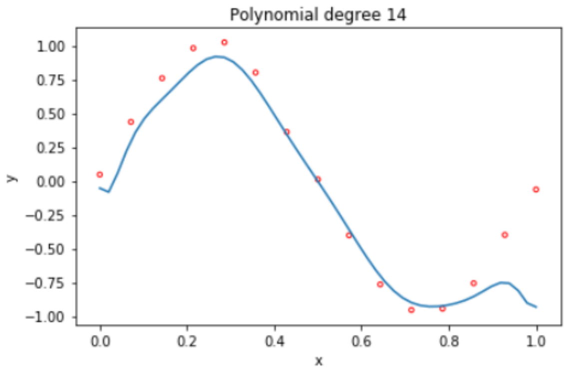
	training	**valid**
1	Degree=10, RMSE=0.0075509741407219465  	Degree=10, RMSE=0.052257899531492416 

2	Degree=10, RMSE=0.005509561461011673 	Degree=10, RMSE=0.47734765454801276 
3	Degree=10, RMSE=0.016849982581915942 	Degree=10, RMSE=0.5990049713882291 
4	Degree=10, RMSE=0.01966295027334867 	Degree=10, RMSE=1.7085018100298333 
5	Degree=10, RMSE=0.009822547482078545 	Degree=10, RMSE=0.7537526978417897 

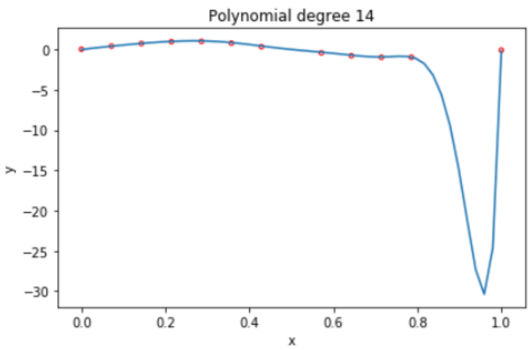
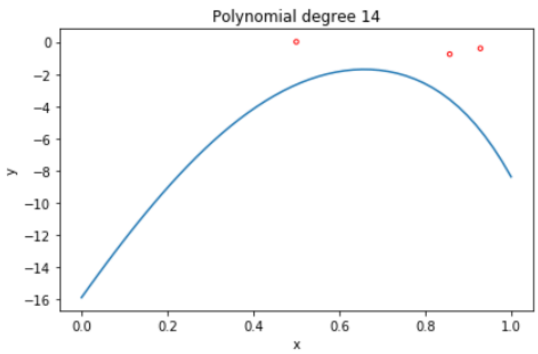
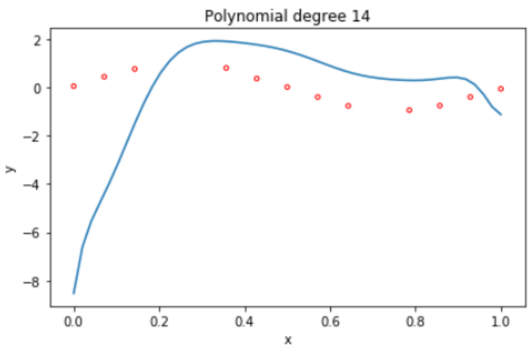
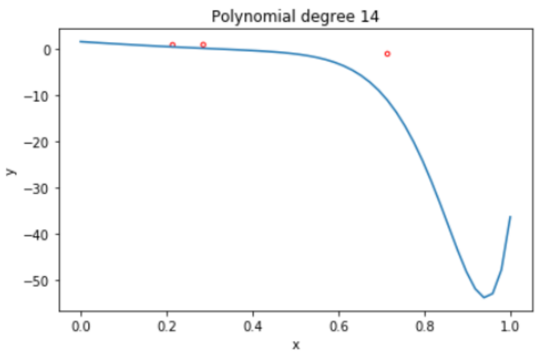
- Degree=14

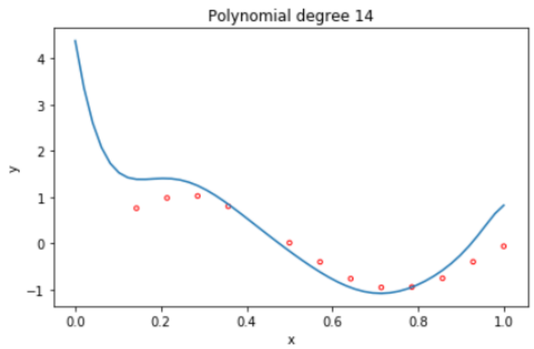
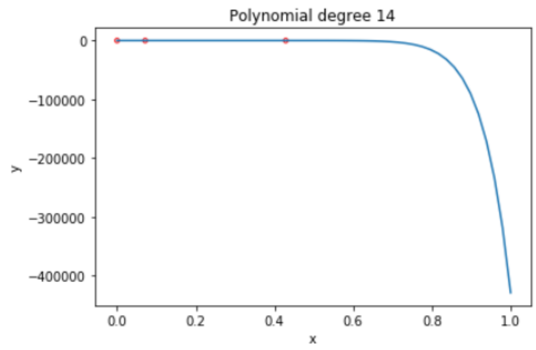
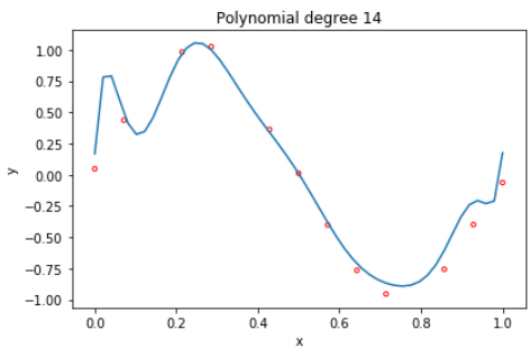
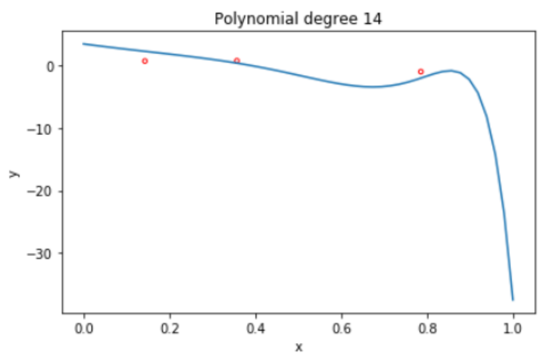
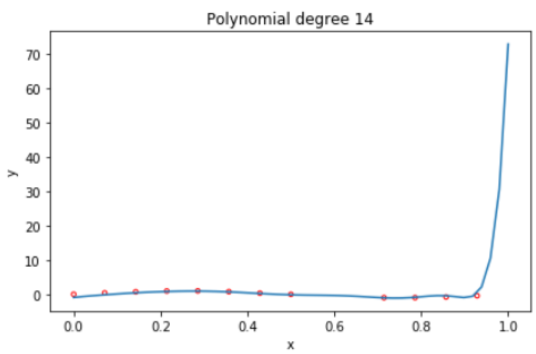
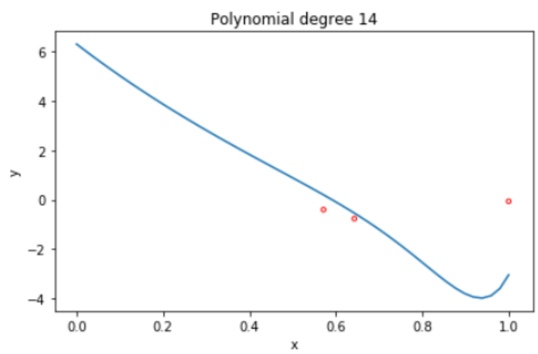
Training error

```
Degree=14, RMSE=0.18215695742583007
```



Five-fold cross-validation errors

	training	**valid**
1	Degree=14, RMSE=0.04286383164989321 	Degree=14, RMSE=2.612141125753481 
2	Degree=14, RMSE=2.196955962159729 	Degree=14, RMSE=4.124748085494175 

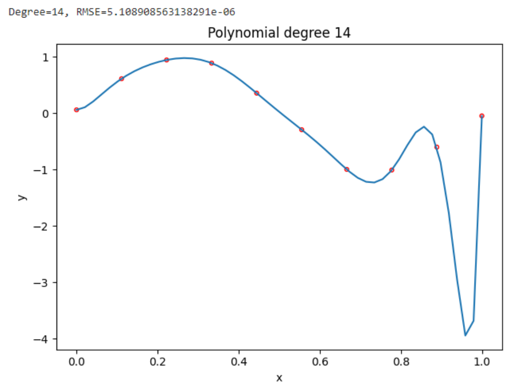
3	Degree=14, RMSE=0.2716989909273514 	Degree=14, RMSE=0.2717773294844128 
4	Degree=14, RMSE=0.07744412576878784 	Degree=14, RMSE=0.7776997101302409 
5	Degree=14, RMSE=0.2939902916102229 	Degree=14, RMSE=1.242799353724011 

- 比較: linear regression 對曲線類型的 dataset(Ex: sin)不適用，應用 polynomial regression 解決曲線類型的 dataset。

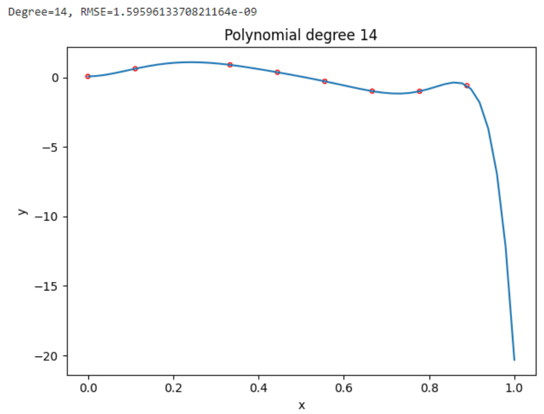
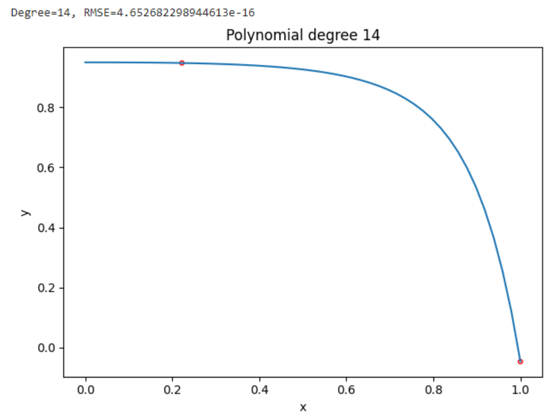
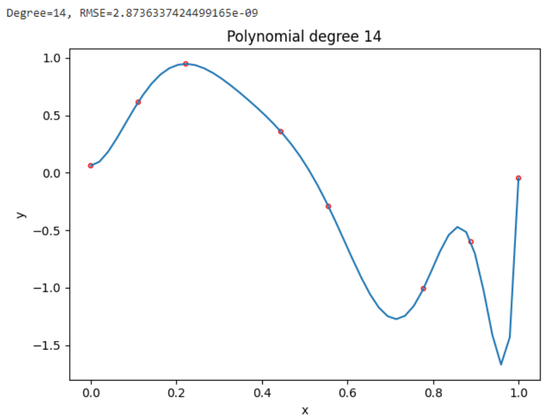
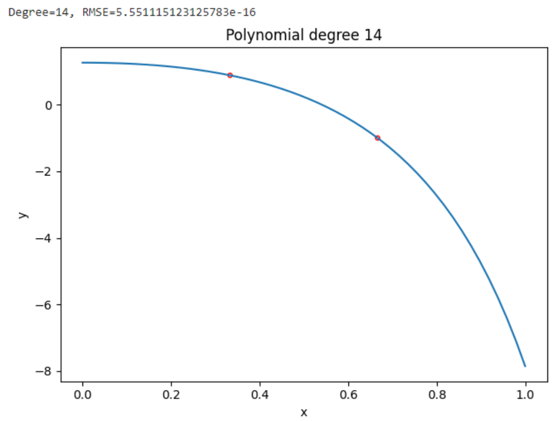
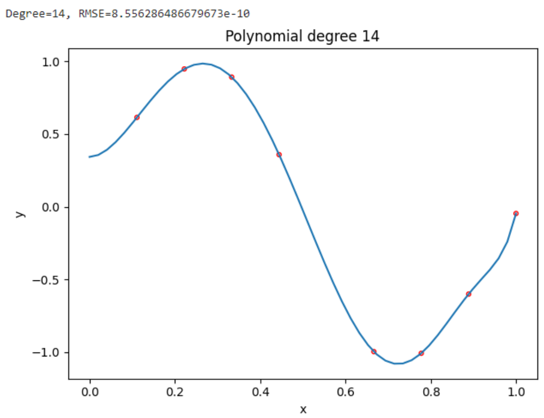
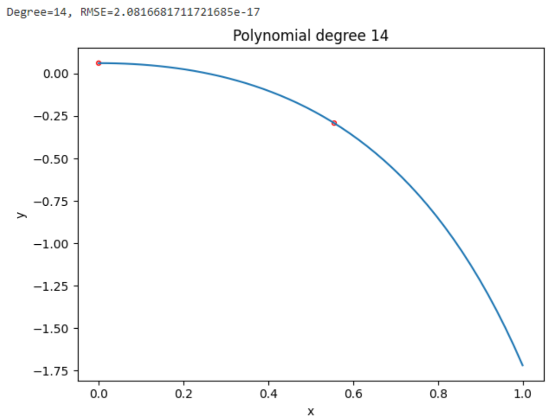
➢ Following 4), perform polynomial regression with degree 14 by varying the number of training data points m = 10, 80, 320. Show the five-fold cross-validation errors and the fitting plots. Compare the results to those in 4).

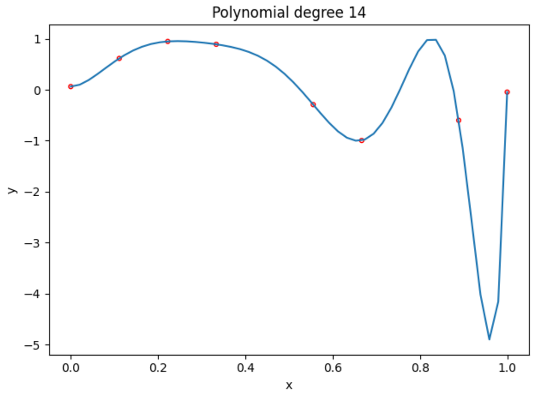
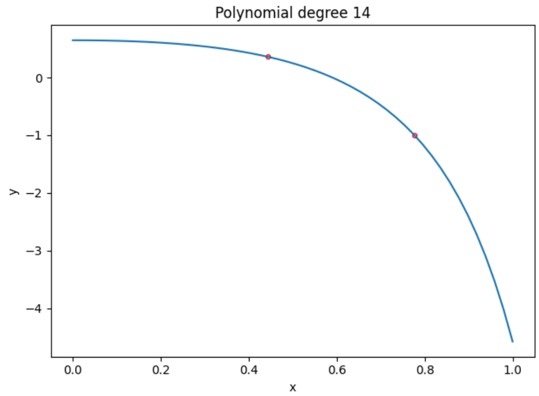
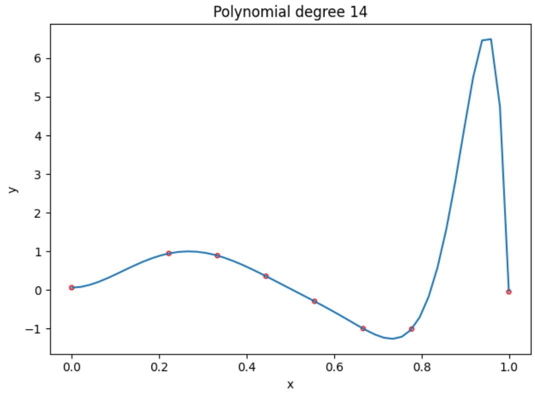
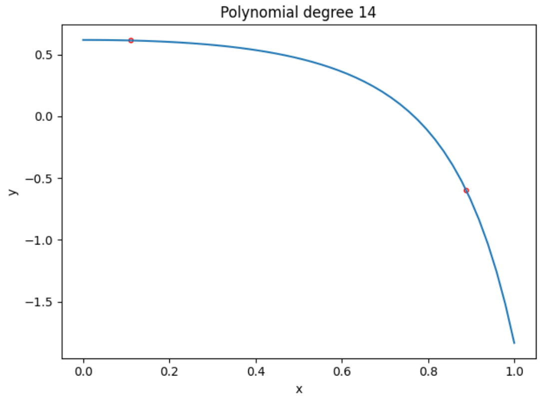
- m=10

  <u>Training error</u>



Five-fold cross-validation errors

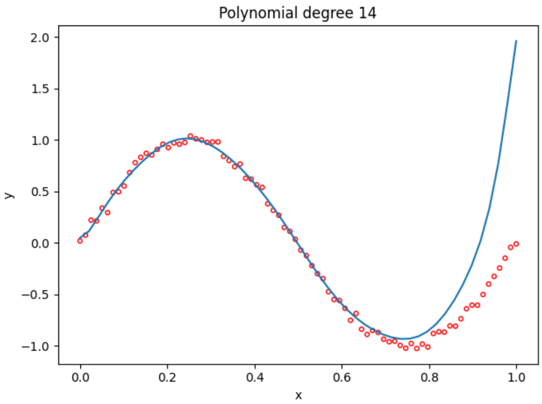
	training	valid
1		
2		
3		

4	Degree=14, RMSE=2.3577491793425743e-09  	Degree=14, RMSE=4.440892098500626e-16 
5	Degree=14, RMSE=9.917788099107515e-09  	Degree=14, RMSE=2.2887833992611187e-16 

- m=80

<u>Training error</u>

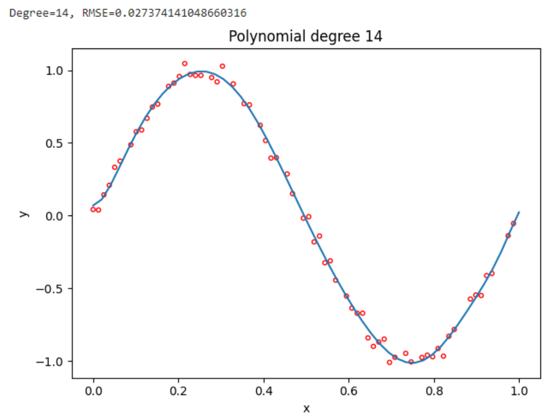
Degree=14, RMSE=0.2750086613851848

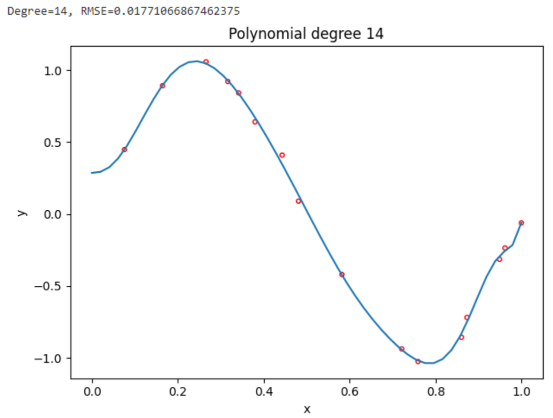


<u>Five-fold cross-validation errors</u>

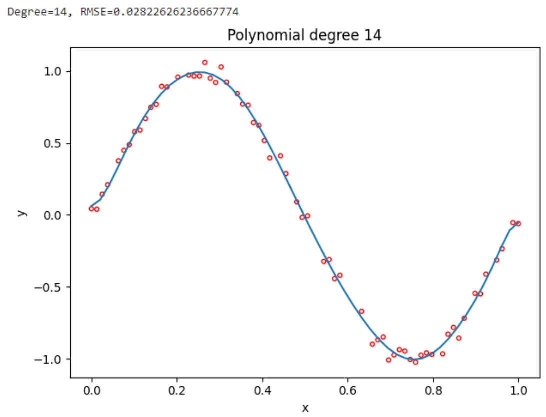
	training	valid

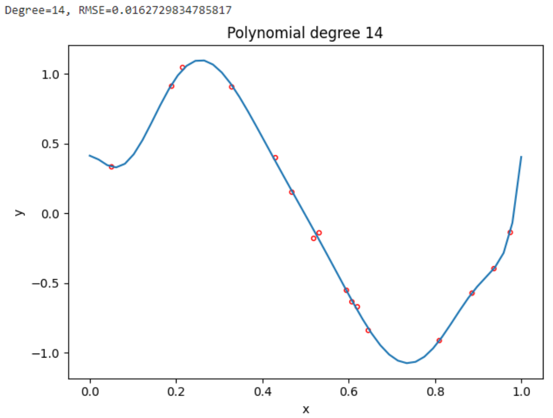
1


Degree=14, RMSE=0.027374141048660316
Polynomial degree 14
y
x

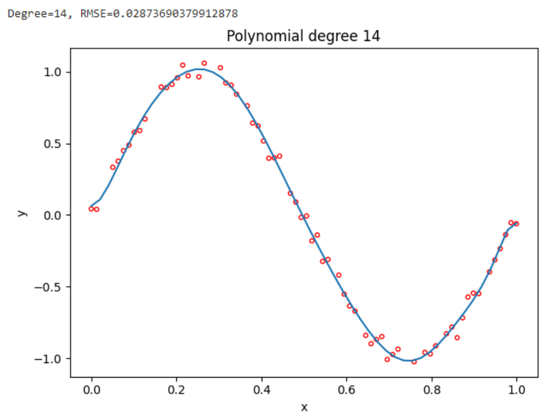

Degree=14, RMSE=0.01771066867462375
Polynomial degree 14
y
x

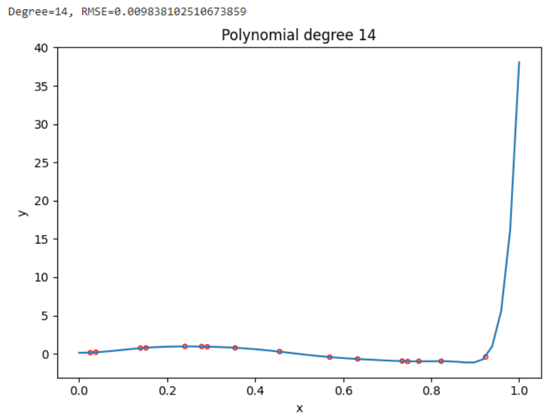
2


Degree=14, RMSE=0.02822626236667774
Polynomial degree 14
y
x

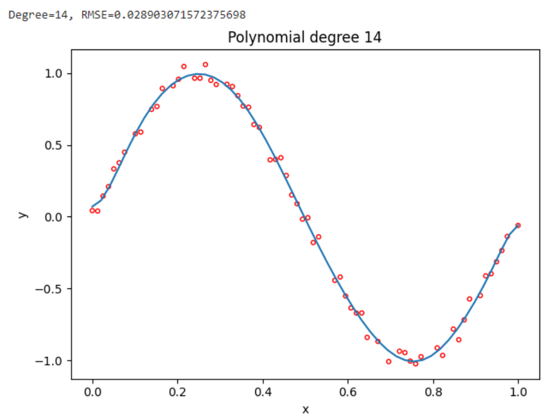

Degree=14, RMSE=0.016272983478581 7
Polynomial degree 14
y
x

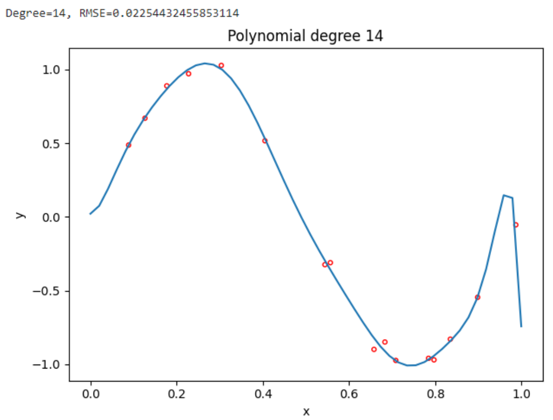
3

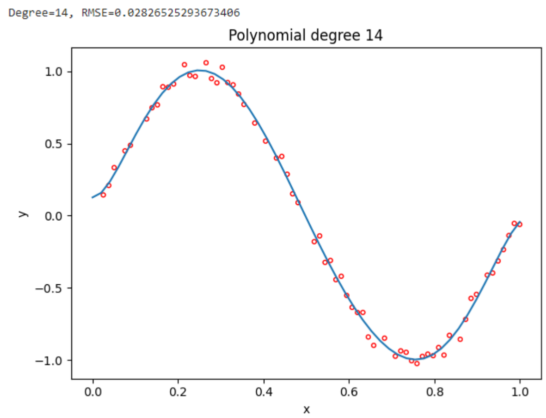
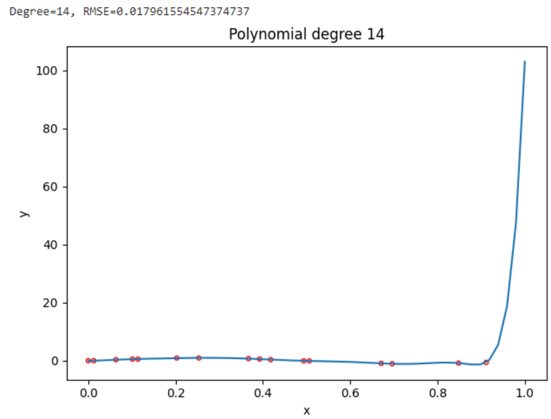

Degree=14, RMSE=0.02873690379912878
Polynomial degree 14
y
x


Degree=14, RMSE=0.009838102510673859
Polynomial degree 14
y
x

4

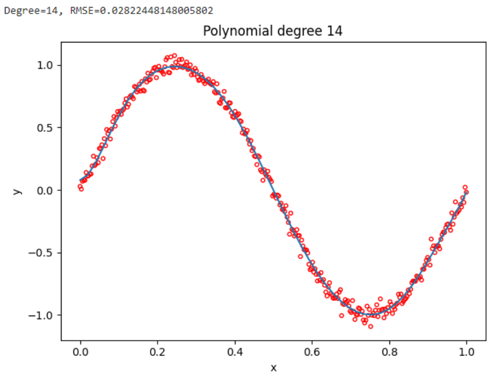

Degree=14, RMSE=0.028903071572375698
Polynomial degree 14
y
x


Degree=14, RMSE=0.02254432455853114
Polynomial degree 14
y
x

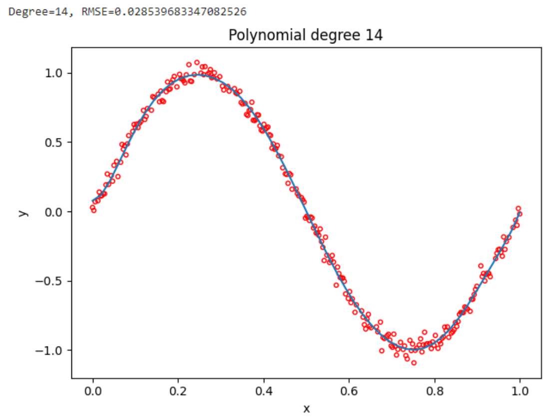
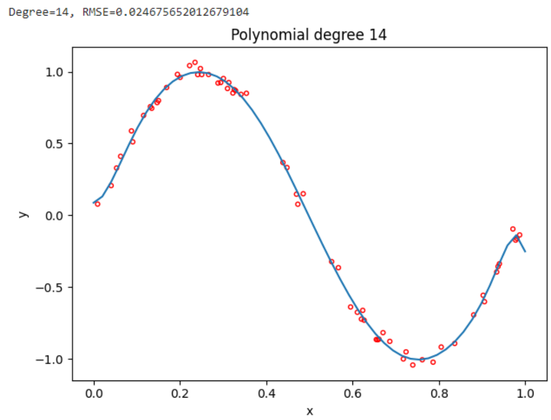
5		

- m=320

<u>Training error</u>

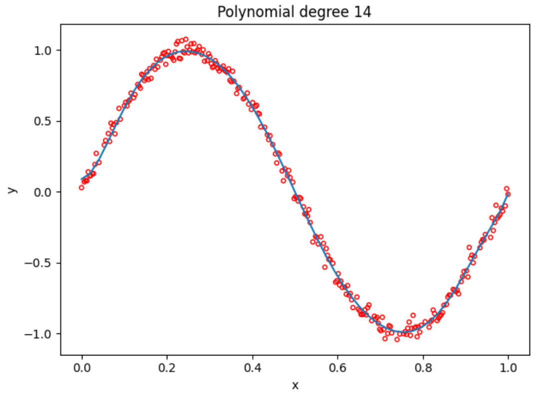


<u>Five-fold cross-validation errors</u>

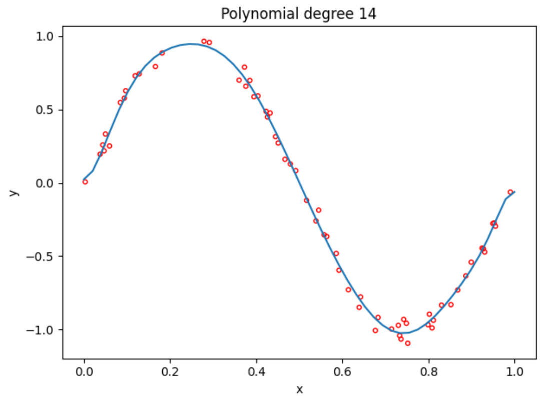
	training	**valid**
1		

2

Degree=14, RMSE=0.02690459989925576

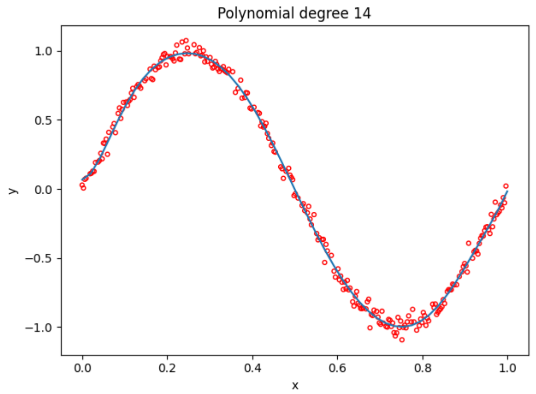


Degree=14, RMSE=0.030052887261067373

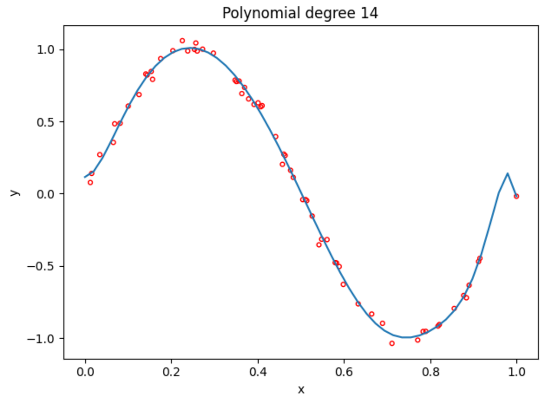


3

Degree=14, RMSE=0.028455162611668746

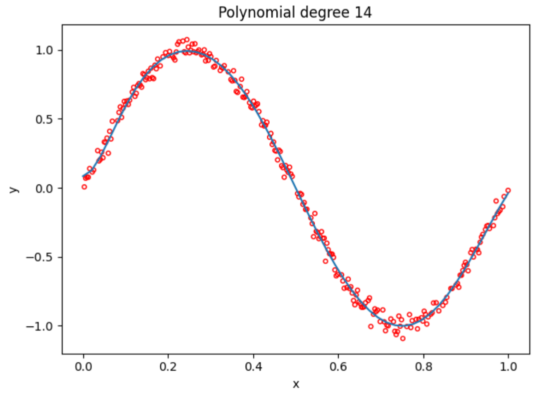


Degree=14, RMSE=0.024655234489205598

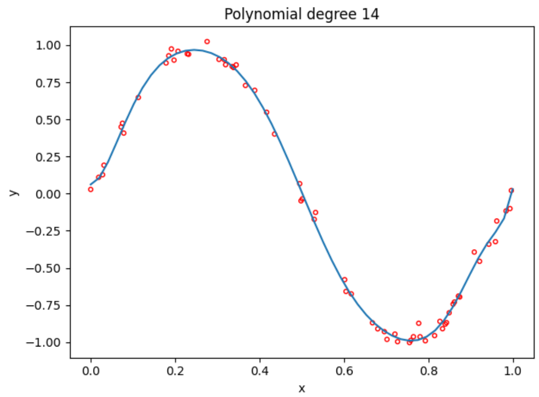


4

Degree=14, RMSE=0.028229440359403838

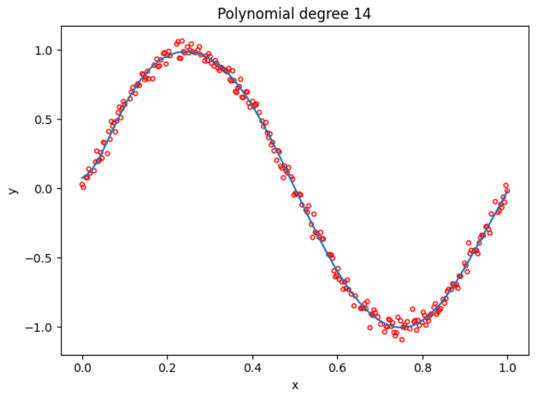


Degree=14, RMSE=0.025973146341301805

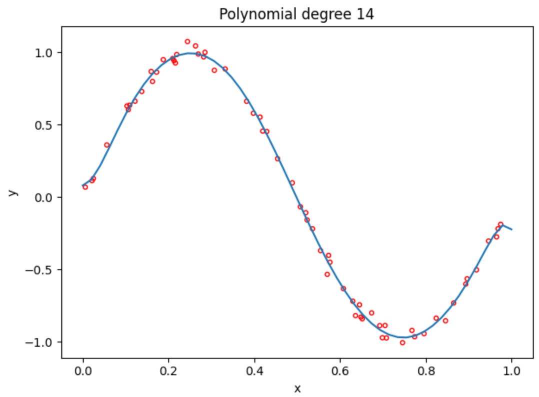


5

Degree=14, RMSE=0.028446494167696772



Degree=14, RMSE=0.024722342500053004



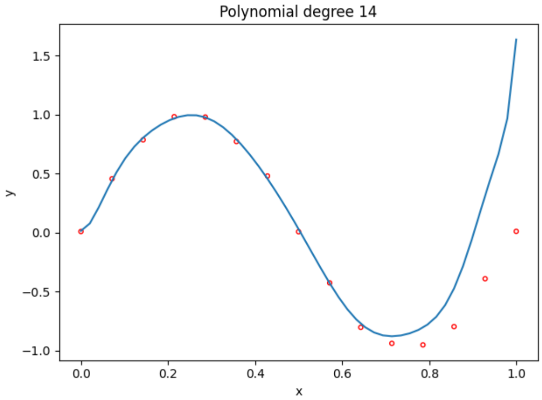
- 比較: m=10 時，因為 data 很少，所以模型遇到沒看過的資料時會隨機猜測，導致預測結果不如預期。對比 m=320 的波形明顯可看出在 num_point 數量多的情況下模型預測準確率較 num_point 少的高很多。

➢ Following 4), perform polynomial regression of degree 14 via regularization. Compare the results by setting λ = 0, 0.001/m , 1/m, 1000/m , where m = 15 is the number of data points (with x = 0, 1/(m−1) , 2/(m−1), . . . , 1). Show the five-fold cross-validation errors and the fitting plots.

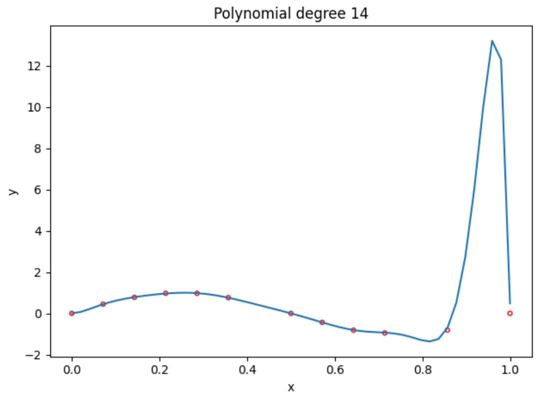
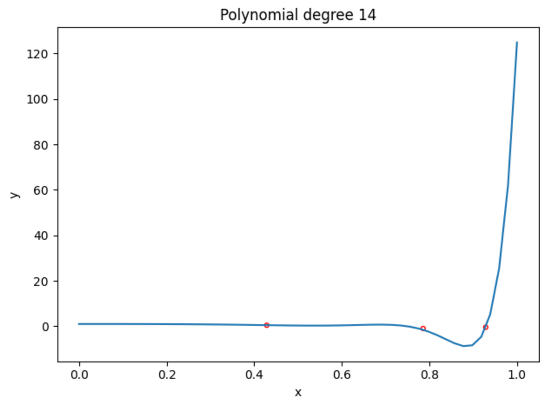
- λ= 0

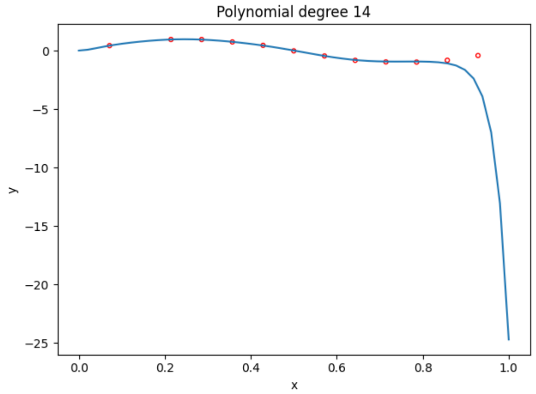
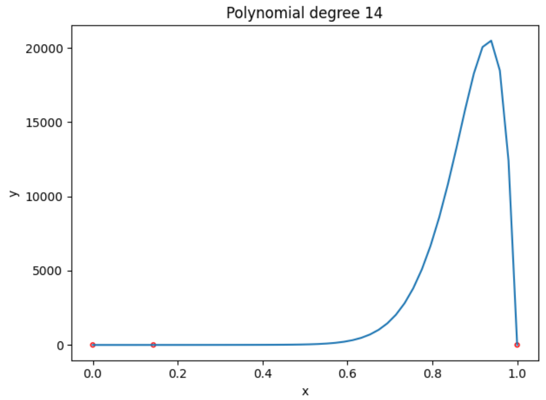
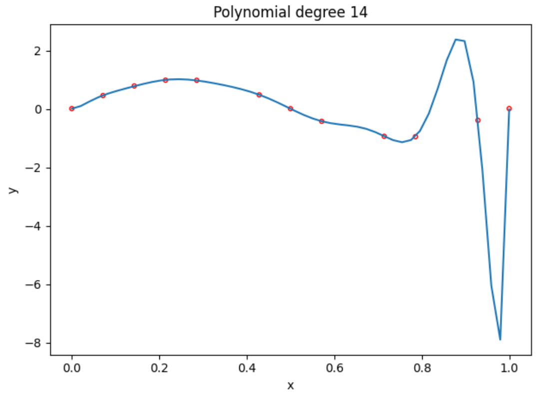
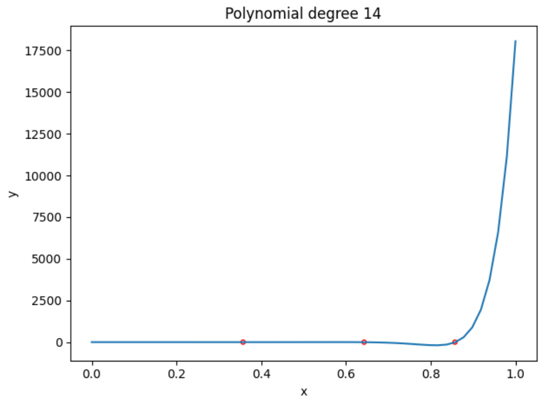
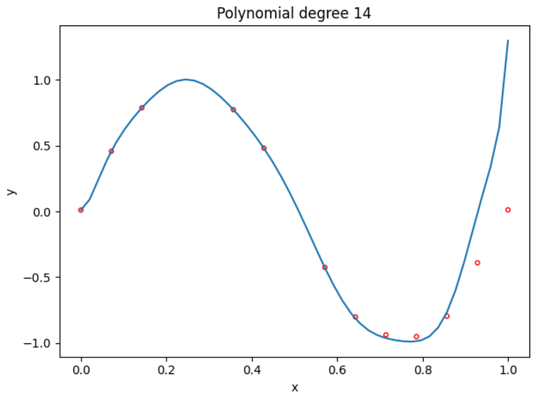
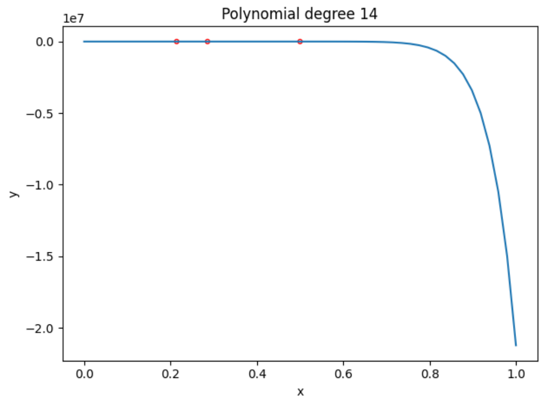
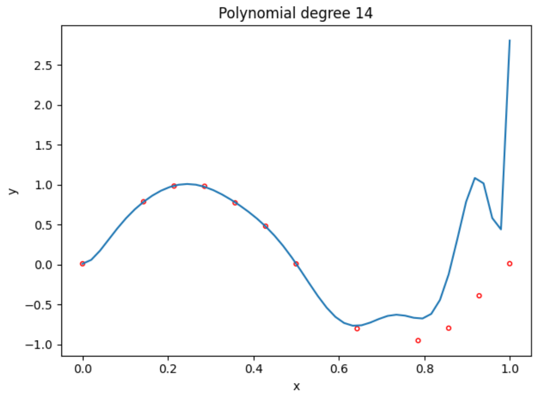
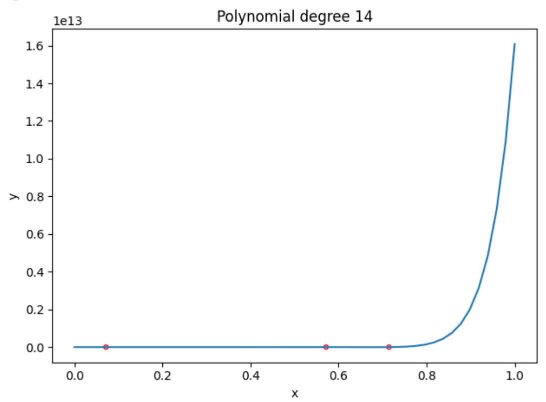
<u>Training error</u>

Degree=14, RMSE=0.3302438202813586



<u>Five-fold cross-validation errors</u>

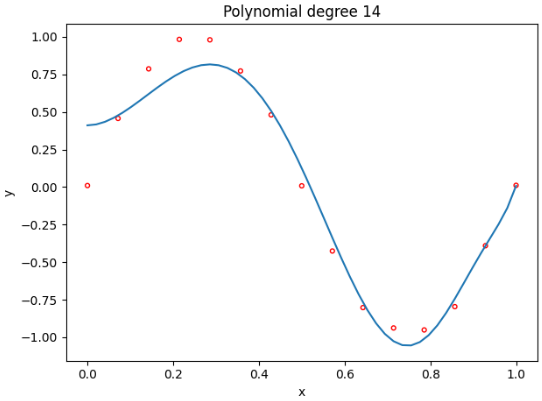
	training	**valid**
1	Degree=14, RMSE=0.10171318410510395 	Degree=14, RMSE=0.2658217545935833 

2	Degree=14, RMSE=0.5340411136129597 	Degree=14, RMSE=0.0833297798658421 
3	Degree=14, RMSE=0.006705815936392849 	Degree=14, RMSE=5.642128015915632 
4	Degree=14, RMSE=0.27368353929368183 	Degree=14, RMSE=0.2366764681273377 
5	Degree=14, RMSE=0.6635628401090699 	Degree=14, RMSE=214.84346220129635 

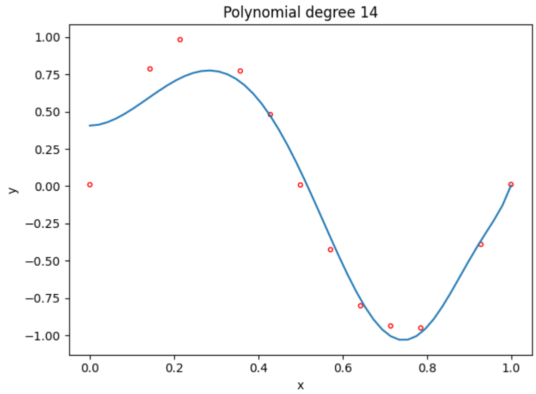
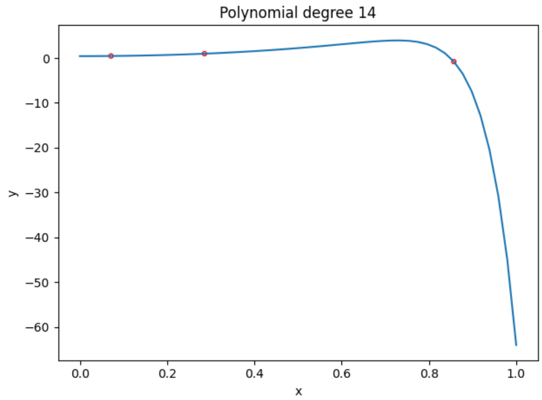
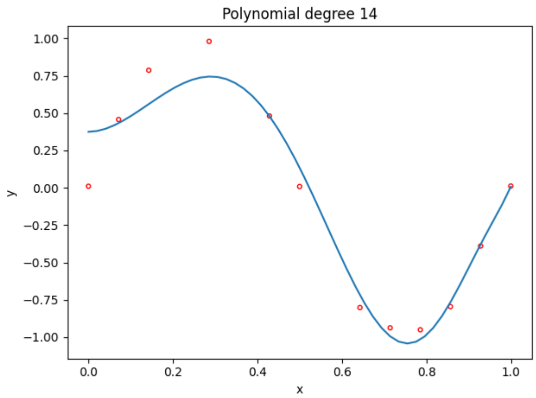
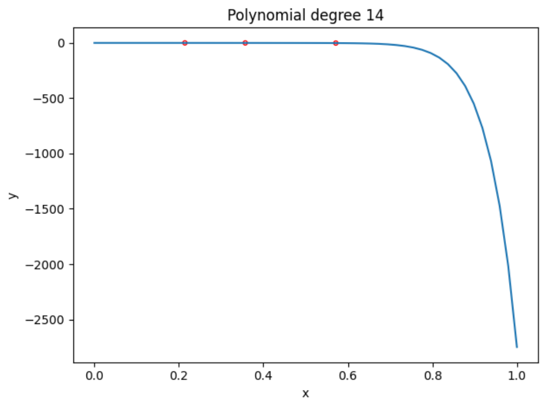
- λ= 0.001/15

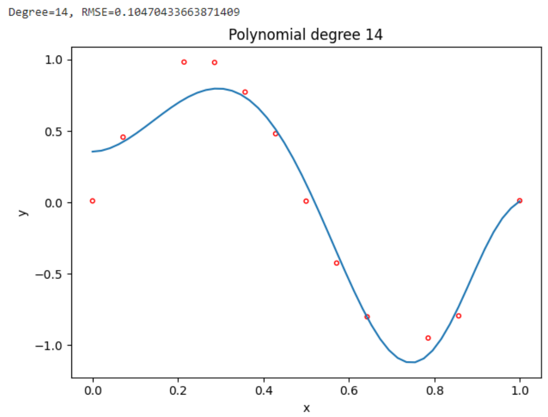
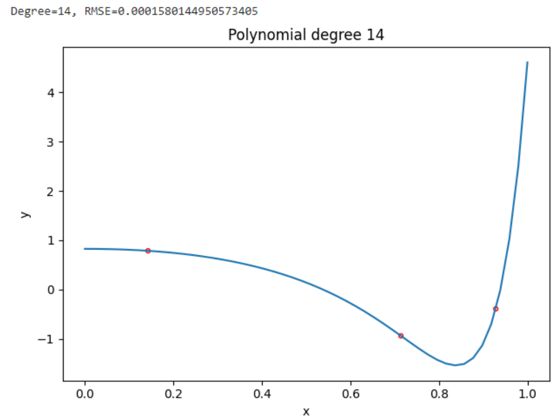
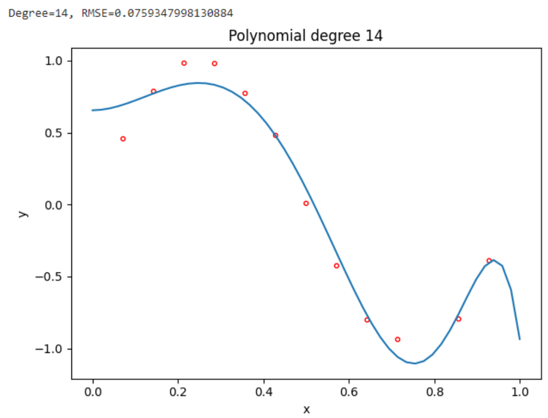
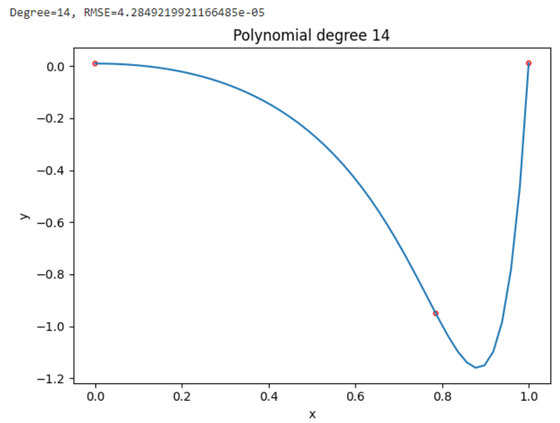
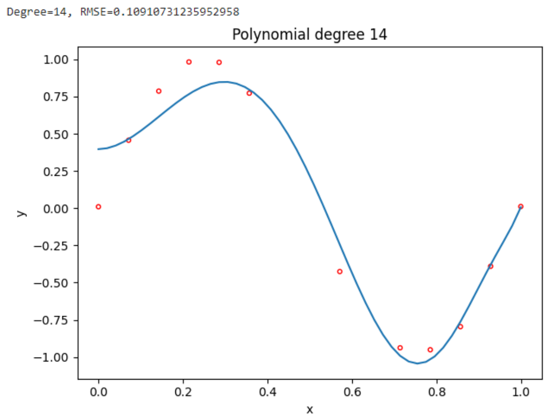
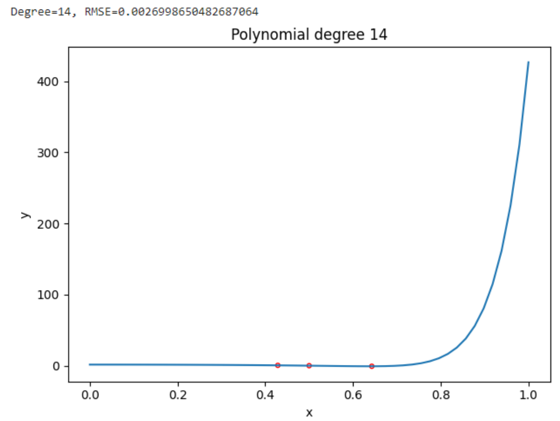
Training error

Degree=14, RMSE=0.100907961947229



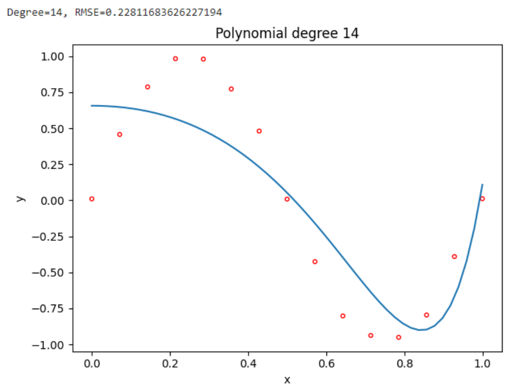
Five-fold cross-validation errors

	training	**valid**
1	Degree=14, RMSE=0.10957180894168682 	Degree=14, RMSE=0.003058046077530753 
2	Degree=14, RMSE=0.10638270483906974 	Degree=14, RMSE=0.001015813918407533 

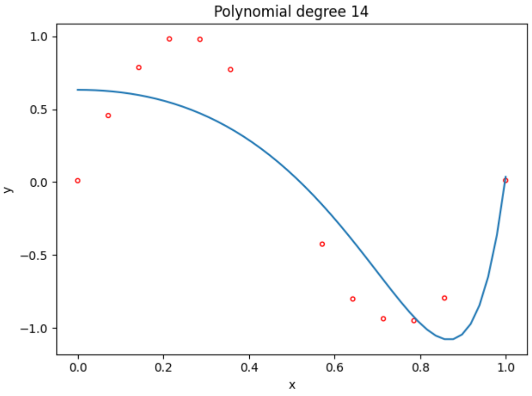
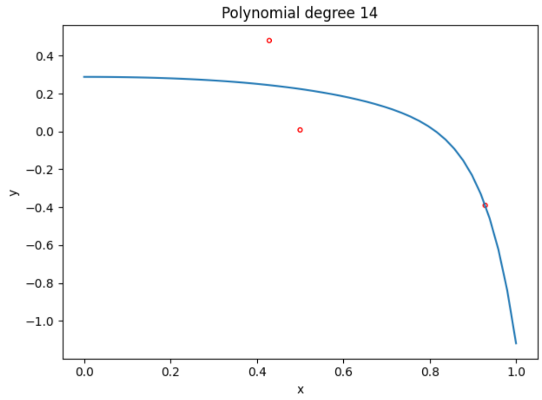
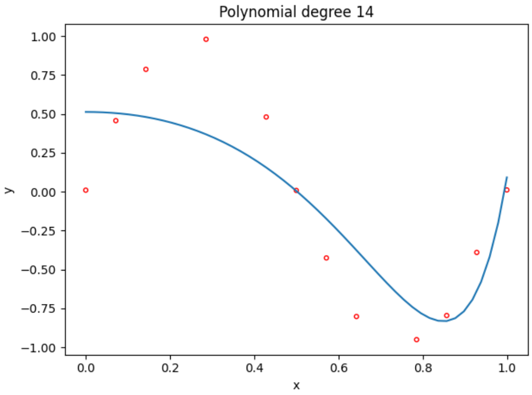
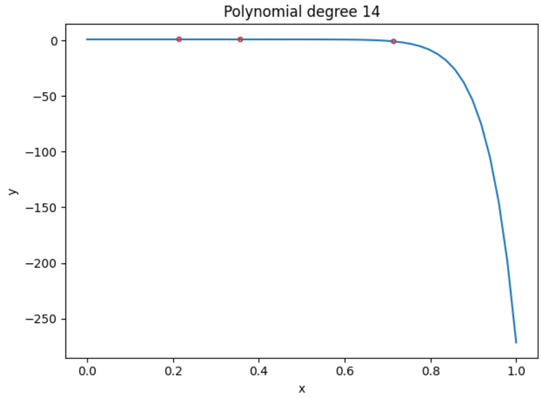
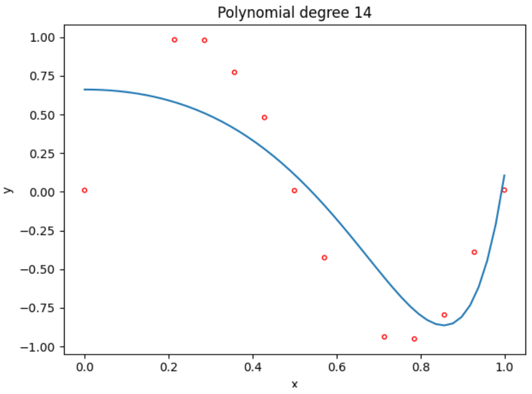
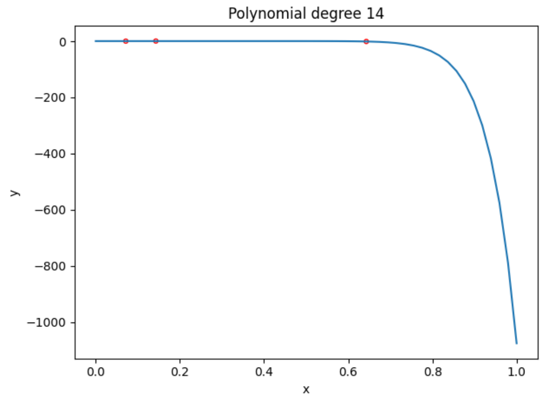
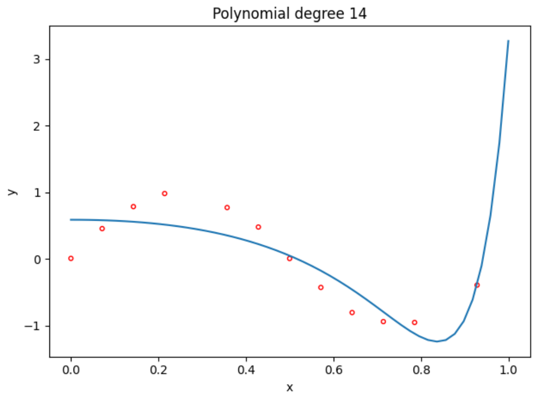
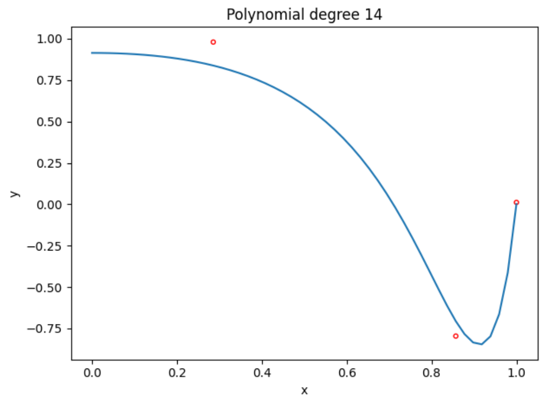
3		
4		
5		

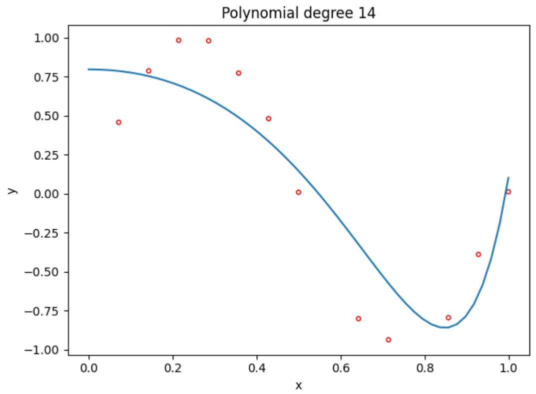
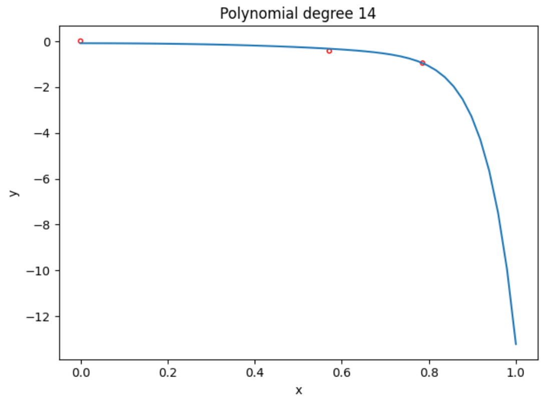
- λ= 1/15

<u>Training error</u>



Five-fold cross-validation errors

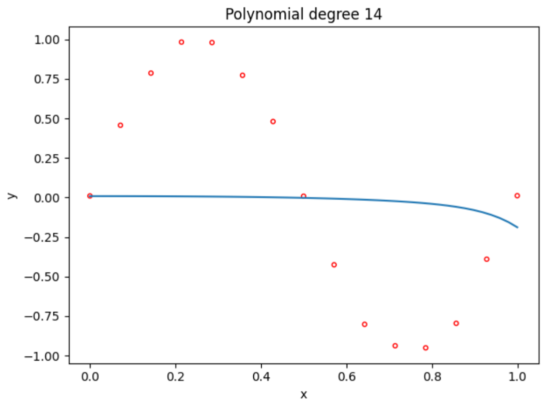
<table>
<tr><th></th><th>training</th><th>valid</th></tr>
<tr><td>1</td><td>Degree=14, RMSE=0.245748467082031
</td><td>Degree=14, RMSE=0.13060864499556255
</td></tr>
<tr><td>2</td><td>Degree=14, RMSE=0.2217018496391975
</td><td>Degree=14, RMSE=0.061317219546042204
</td></tr>
<tr><td>3</td><td>Degree=14, RMSE=0.24037169564131938
</td><td>Degree=14, RMSE=0.09573030141252975
</td></tr>
<tr><td>4</td><td>Degree=14, RMSE=0.21361585322383317
</td><td>Degree=14, RMSE=0.06863846922599011
</td></tr>
</table>

5	Degree=14, RMSE=0.19258320481981725 	Degree=14, RMSE=0.056757140482775505 

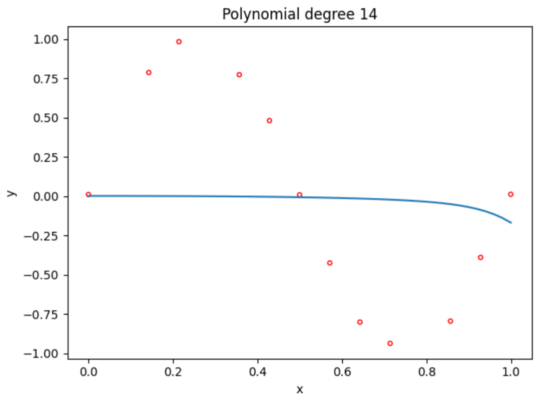
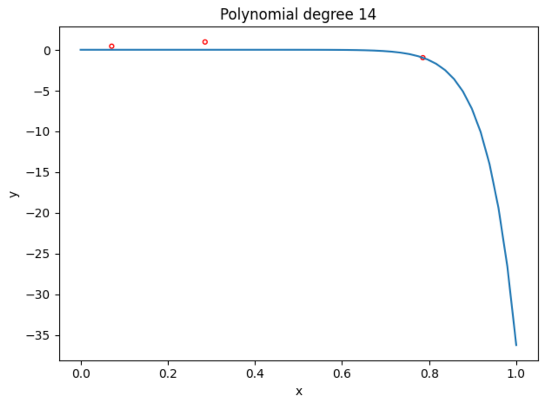
- λ= 1000/15

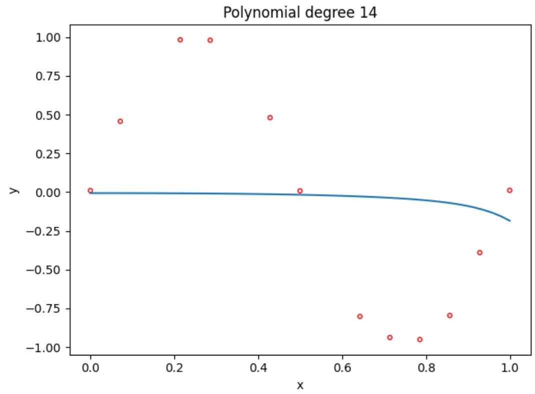
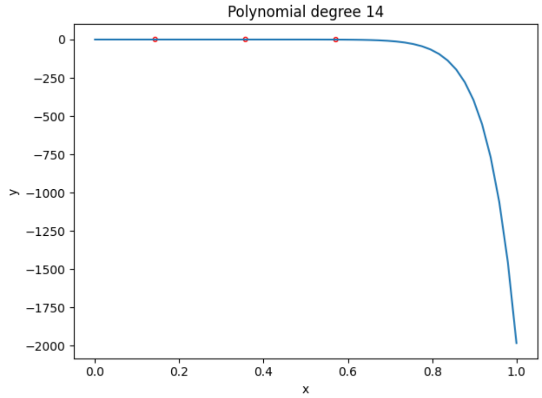
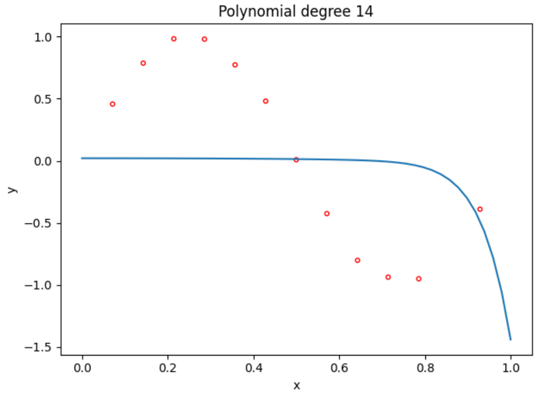
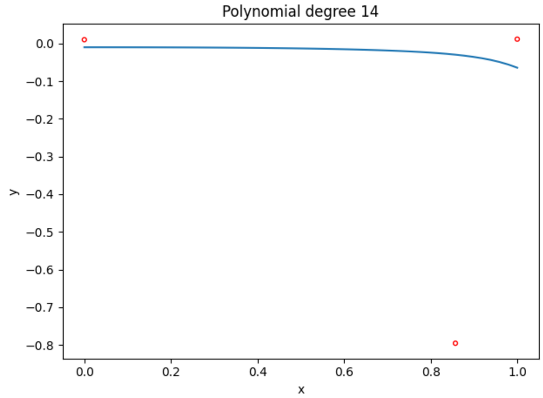
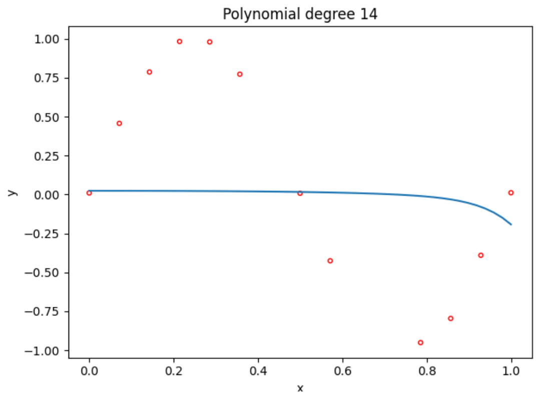
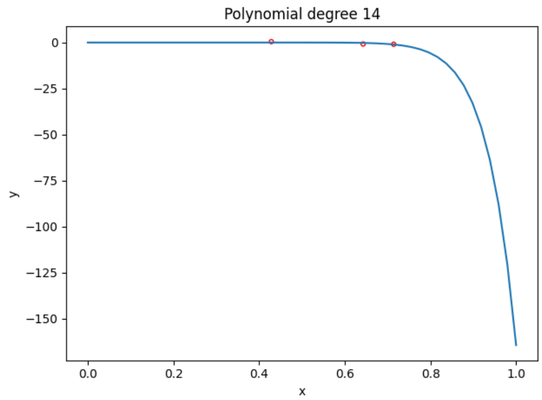
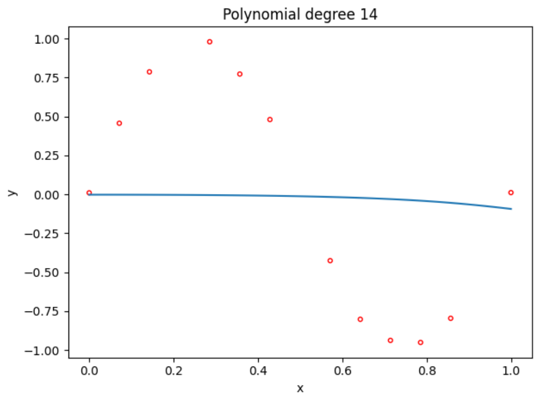
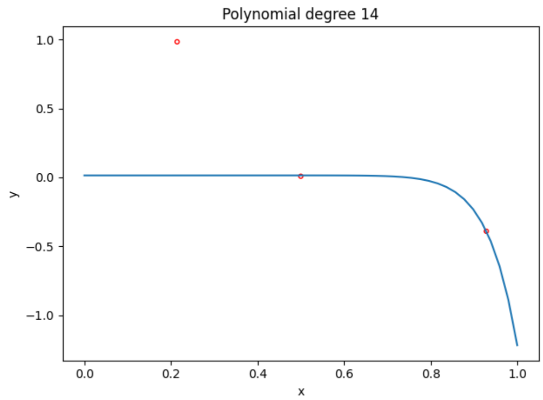
Training error

Degree=14, RMSE=0.4713576802594253

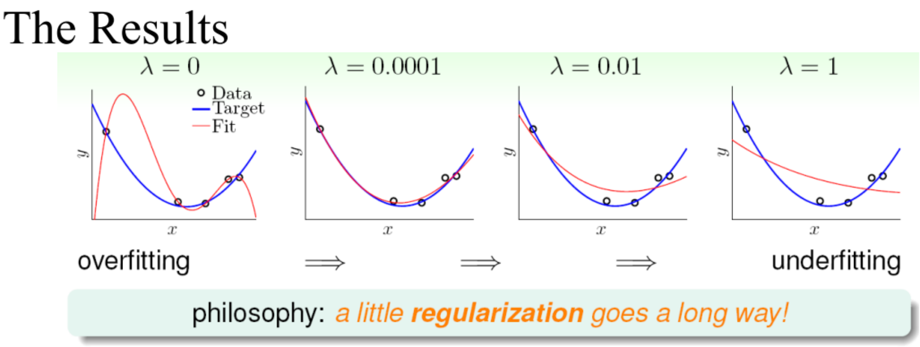


Five-fold cross-validation errors

	training	valid
1	Degree=14, RMSE=0.4438510777454821 	Degree=14, RMSE=0.4300883302368353 

2	Degree=14, RMSE=0.4693961737558944 	Degree=14, RMSE=0.4371590663187278 
3	Degree=14, RMSE=0.49830922753483975 	Degree=14, RMSE=0.31438418511465227 
4	Degree=14, RMSE=0.45536663852178516 	Degree=14, RMSE=0.3127302511377 
5	Degree=14, RMSE=0.4846834158948426 	Degree=14, RMSE=0.39540147527277564 

- Conclusion: The observation from your results.
    - ➢ Linear regression 只適用於線性關係的資料，但當我們的資料是非線性的就必須用 polynomial regression 才能畫出最佳擬合線。龐大的資料集在做 five-fold 時，比較不會因資料量不足找不到逆矩陣而出現錯誤，但過少的資料集也可能造成訓練資料不足，因而導致 overfitting。
    - ➢ 改變λ值， λ=0 時模型無法彈性的做出判斷 (模型複雜度太高)，由擬合圖可看出有 overfitting 的問題; λ=1000/m 時彈性過大因此模型無法判斷所學習的資料，由擬合圖看出有 underfitting 的問題; λ=0.001/m 最為精準，能畫出最佳擬合線。



實驗結果完全符合課堂簡報中的結果。

- Discussion: The questions or the difficulties you met during the implementation.

原本以為是因為點太少而一直出現找不到逆矩陣的問題

```
     86
     87 def _raise_linalgerror_singular(err, flag):
---> 88     raise LinAlgError("Singular matrix")
     89
     90 def _raise_linalgerror_nonposdef(err, flag):

LinAlgError: Singular matrix
```

將 np.linalg.inv 改成 np.linalg.pinv。

參考資料: https://blog.csdn.net/weixin_43977640/article/details/109908976