Lab 09

Name: 白騏瑞

Student ID: B8611027

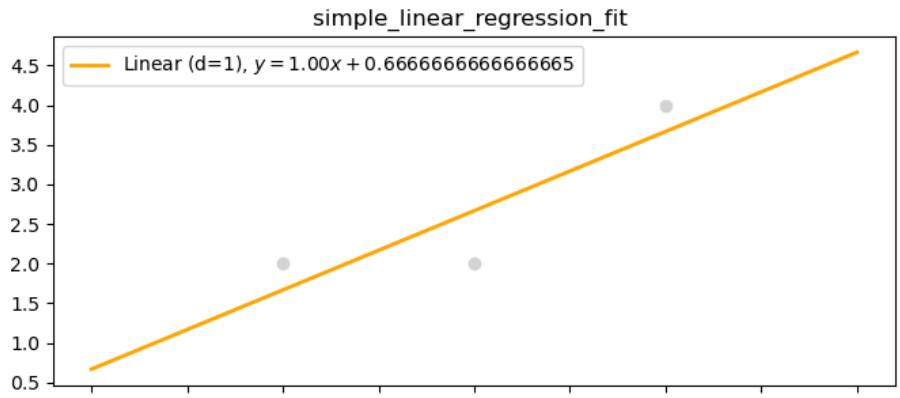
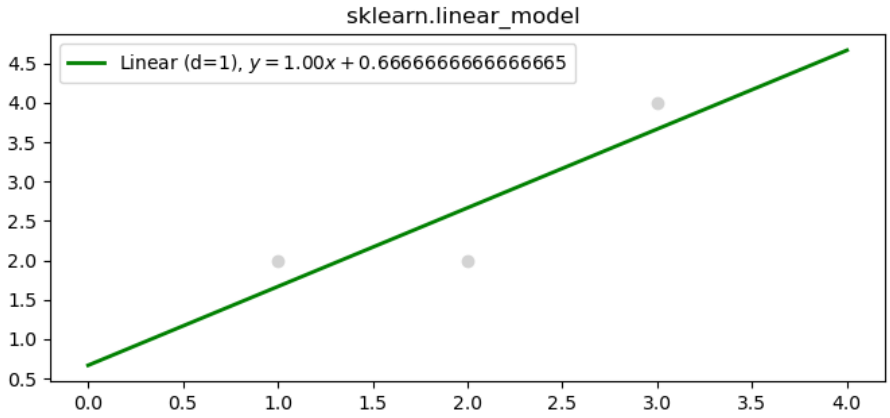
Total Score:

A. Multiple Choice (35 points, 5 points each question)

#	Answer	Explanation (Optional)	Score
1	A		
2	C		
3	B	(a) 0.4 (b) 0.8	
4	B	Increased complexity of the underlying model may increase the overfitting problem. Decreasing the complexity may help in reducing the overfitting problem. Noise in the training data can increase the possibility for overfitting. Noise reduction can help in reducing the overfitting.	
5	C		
6	C		
7	E		

B. Simple Linear Regression (30 points, 10 points each question)

#	Description	Score
1	Scatter 在下面兩個圖中有出現(灰色的點)	
2	$\beta 0$：0.67 $\beta 1$：1	

		
3	The best-fit line is same as the one in question 2.，因為都是用最小平方法求出來的。 	

C. Multiple and Polynomial Regression (35 points)

#	Description	Score

1	(2)
	• Train: 80%, Test: 20%
	`mse= 33.198187930253276`
	• Train: 67%, Test: 33%
	`mse= 31.80091968806165`
	• Train: 50%, Test: 50%
	`mse= 44.13102124913964`
	發現 training data 越大 ，test 的最小平方差越小。
	上網查發現大部分的人會選擇 Train：80%, test：20%，因此最終決定選擇這個比例。	

2	1)不確定是不是資料有點多筆，所以 mse 的數值有點大。	

	2) 1. Age 與 age^2 的係數分別為： 5.63042235e+00 與 5.19351275e+00 2. log2(page_views)的係數為： -9.94646158e-02 3. big_club 的係數則是： 5.30309533e+00 3) 可以嘗試用運用更多的係數去調整 model。 4) 嘗試過各種數值，發現 training data 越大 ，test 的最小平方差越小，所以我認為應該還沒有 overfit	