**Linear regression**

(一) 單變數迴歸

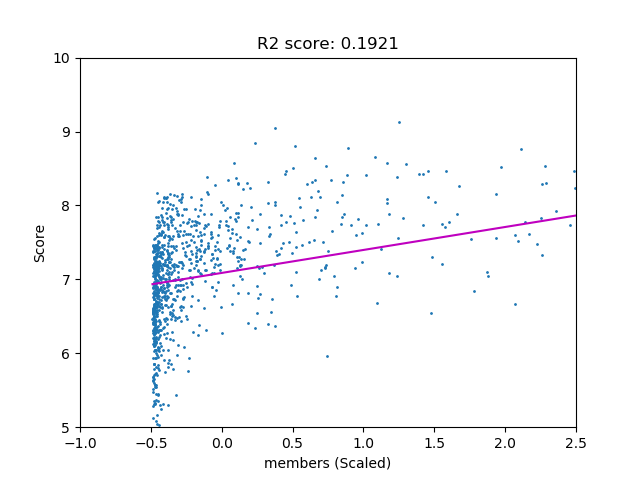
針對members、popularity、favorites、studio\_num、episodes，這些動畫開播之前就可以取得的5個離散型變數，分別以他們五個作為單變數線性迴歸的自變數，來預測動畫的評分。

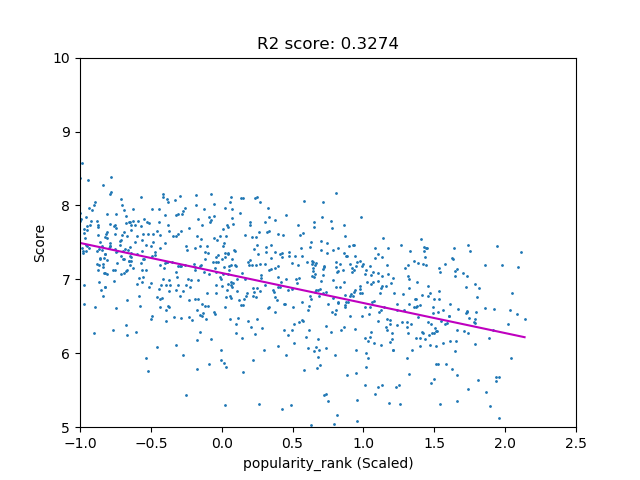
在做迴歸分析之前，先以下述的標準篩掉資料中的離群值：

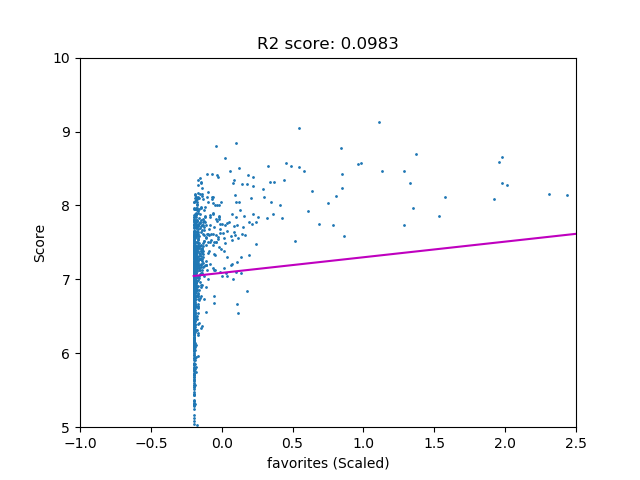
* Drop scoreBy <= 1000
* n\_data = 5007 (i.e. 共有5007筆資料)
* 80% of training set, 20% of testing set

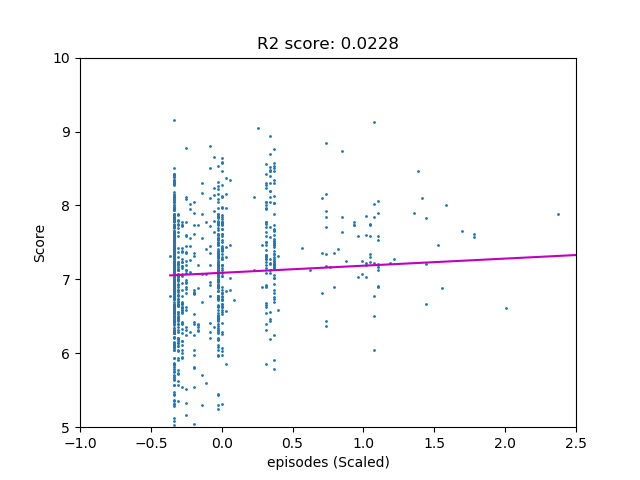
以下是單變數線性迴歸分析的結果：

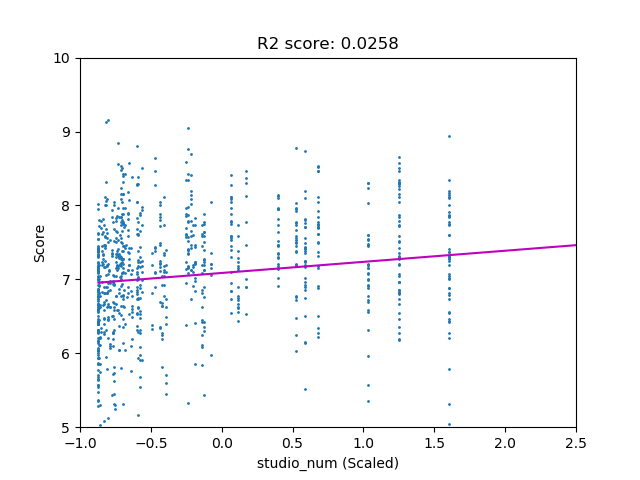
(使用檔案：SingleVariable\_Regression.py)











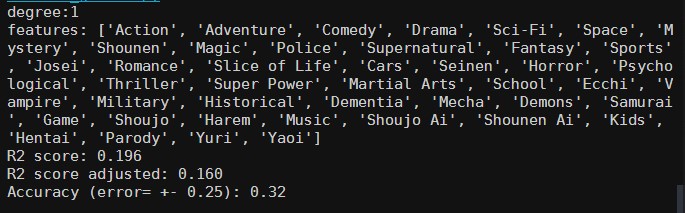
可以看的出來其中members和popularity，對於預測評分來說，是較為重要的自變數。

(二) 多變數迴歸

使用檔案：SingleVariable\_Regression.py

這裡我們針對genres ，也就是 [Action, …, Yuri]這43個binary variable，以他們為自變數，作多變數線性迴歸來預測動畫評分。

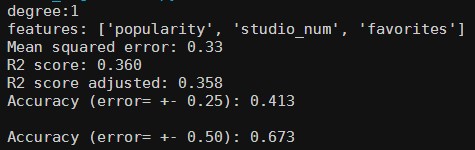
以下是預測結果：(使用檔案：MultiVariable\_Regression\_genres.py)



大約可獲得三成的準確度。

另外，針對第一部分單變數線性迴歸，經過窮舉各種自變數的組合，我找出以[studio\_num, popularity, favorites]這三個變數做為自變數的話，可以獲得最高的準確度，以下是預測結果：

(使用檔案：MultiVariable\_Regression.py)

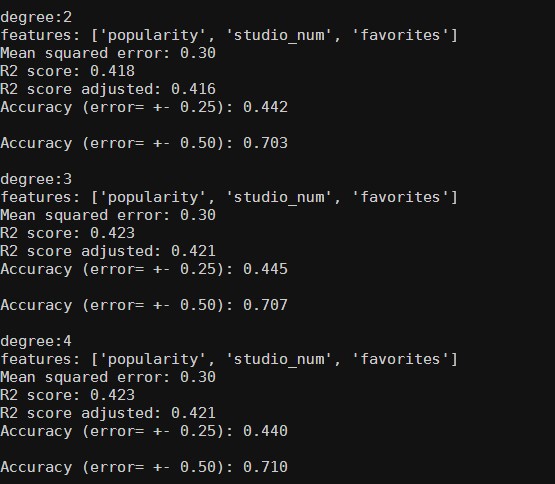


大約有41%的準確度，若是放寬誤差到正負0.5分，則可來到67%的準確度

**Polynomial regression**

針對前面做過多變數線性迴歸的 [studio\_num, popularity, favorites]，以他們自變數，分別做二、三、四次方的多變數迴歸來預測動畫評分。

以下是預測結果：(使用檔案：MultiVariable\_Regression.py)



|  |  |  |  |
| --- | --- | --- | --- |
| Polynomial degree | 2 | 3 | 4 |
| **R2\_score** | **0.418** | **0.423** | **0.423** |
| R2\_score\_adjusted | 0.416 | 0.421 | 0.421 |
| Accuracy (±0.25) | 0.442 | 0.445 | 0.440 |
| Accuracy (±0.5) | 0.703 | 0.707 | 0.710 |

可以看到約有44%的準確度，在誤差為正負0.5分時，準確度可高達70%

另外也可以觀察到，除了從一次方提高至二次方時，R^2\_Score和accuracy有略微提高之外，在二次方以上的狀況時，提高迴歸式的次方，對預測的準確度以及R^2\_Score並沒有太大的幫助。