**Lab Regression 2 (22/4/2022)**

ລະຫັດນັກສຶກສາ :**205Q001019**

ຊື່ ແລະ ນາມສະກຸນ:**ທ້າວ ນູຊົ່ວ ເຮີ 3CW1**

ຈົ່ງຕອບຄຳາຖາມຕໍ່ນີ້ໃຫສໍາເລັດດ້ວຍການນຳໃຊ້ຄໍາສັ່ງຂອງ Python:

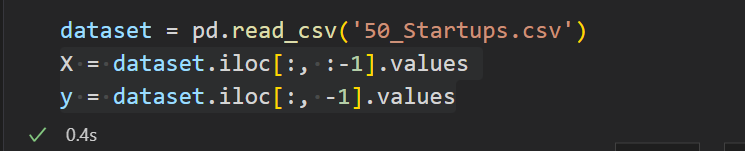
1. Multiple Linear Regression

1.1. ຈາກຊຸດຂໍ້ມູນ (Data Set) 50\_Startups.csv, ຈົ່ງບອກຈຳນນວນຖັນ , ແຖວ (shape) ແລະ ເພີ່ຂໍ້ມູນໃສ່ຕາຕະລາງໃຫ້ສຳເລັດ :

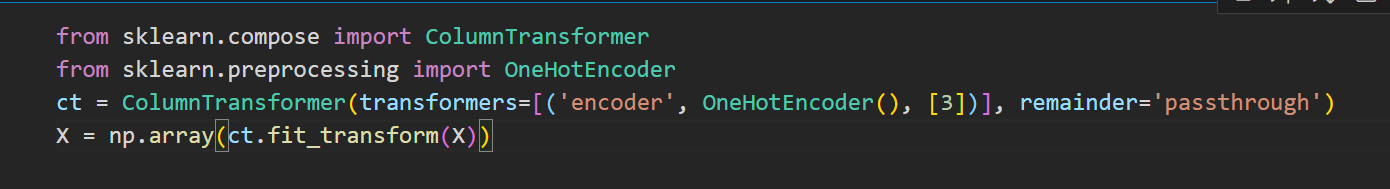
|  |  |
| --- | --- |
| ຖັນ | 5 |
| ແຖວ | 50 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | R&D Spend | Administration | Marketing Spend | State | Profit |
| 0 | 165349.20 | 136897.80 | 471784.10 | New York | 192261.83 |
| 1 | 162597.70 | 151377.59 | 443898.53 | California | 191792.06 |
| 2 | 153441.51 | 101145.55 | 407934.54 | Florida | 191050.39 |

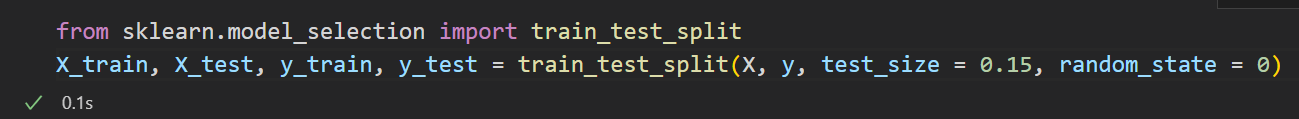
1.2. ຈົ່ງກຳນົດຕົວປ່ຽນເອກະລາດ (Independent Variables X) ໃຫ້ເປັນ R&D Spend, Administration,  Marketing Spend ແລະ State. ກຳນົດຕົວປ່ຽນຕາມ (Dependent Variables y) ໃຫ້ເປັນ Profit.



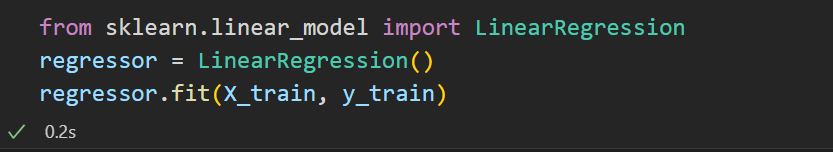
1.3. ຈົ່ງທຳການຕຽມຊຸດຂໍ້ມູນ (Preprocessing) ດ້ວຍການເຮັດ OneHotEncoder ຂອງຕົວປ່ຽນອິດສະຫຼະ X.



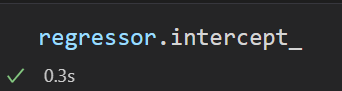
1.4 ຫຼັງຈາກຕຽມຊຸດຂໍ້ມູນສຳເລັດ, ຈົ່ງຂຽນຄຳສັ່ງເພື່ອແຍກຊຸດຂໍ້ມູນເປັນສອງພາກສ່ວນຄື: ຊຸດຮຽນ 85% ແລະ ຊຸດທົດສອ 15%?



1.5. ຈົ່ງສ້າງໂມເດວ Linear Regression ແລະ ທຳາການປະມວນຜົນ (fit) ຊຸດຂໍ້ມູນຮຽນຈາກຂໍ້ 1.4



1.6. ຈົ່ງຄິດໄລ່ຄ່າ intercept\_ ແລະ coef\_ ຂອງໂມເດວຫຼັງການປະມວນຜົນ (fit).

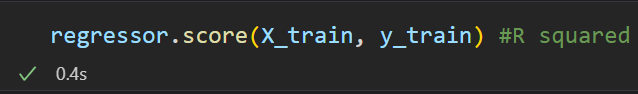


42190.78496234339



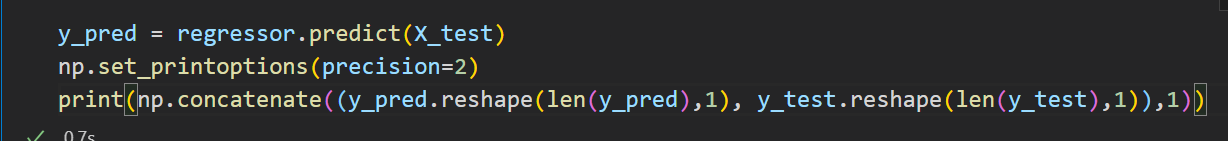
array([ 2.12734825e+02, -1.14601240e+03, 9.33277580e+02, 7.73051855e-01, 3.50926156e-02, 3.60827279e-02])

1.7. ຈົ່ງຄິດໄລ່ຄ່າ R-squared ຂອງໂມເດວຫຼັງການປະມວນຜົນ (fit) ຊຸດຂໍ້ມູນຮຽນຈາກ ຂໍ້1.4

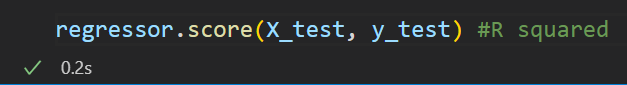


0.9522700739455511

1.8. ຈົ່ງທົດສອບໂມເດວຂອງຊຸດຂໍ້ມູນ 15% ຈາກຂໍ້ 1.4.



1.9. ຈົ່ງຄິດໄລ່ຄ່າ R-squared ຂອງໂມເດວຫຼັງການທົດສອບຊຸດຂໍ້ມູນທົດສອບຈາກຂໍ້1.8.



0.9190315130435046

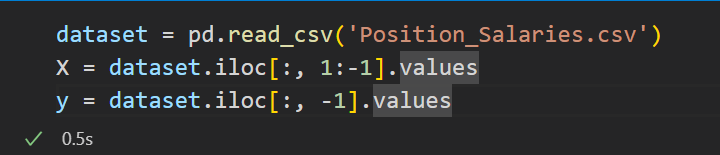
2. Polynomial Regression

2.1 ຈາກຊຸດຂໍ້ມູນ (Data Set) Position\_salaries.csv, ຈົ່ງບອກຈຳນວນຖັນ, ແຖວ (shape) ແລະ ເພີ່ມຂໍ້ມູນໃສ່ຕາຕະລາງໃຫ່ສຳເລັດ: 

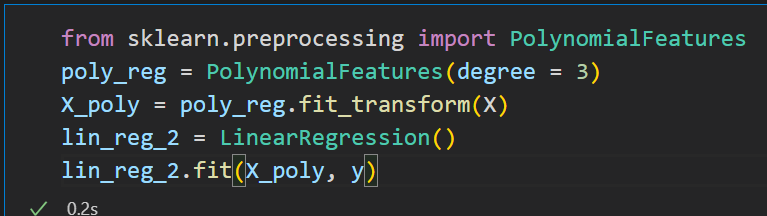
|  |  |  |  |
| --- | --- | --- | --- |
|  | Position | Level | Salary |
| 0 | Business Analyst | 1 | 45000 |
| 1 | Junior Consultant | 2 | 50000 |
| 2 | Senior Consultant | 3 | 60000 |

|  |  |
| --- | --- |
| ແຖວ | 10 |
| ຖັນ | 3 |

2.2 ຈົ່ງກຳນົດ Level ໃຫ້ເປັນຕົວປ່ຽນເອກະລາດ (Independent Variables X). ກຳນົດ Salary ໃຫ້ເປັນຕົວປ່ຽນຕາມ (Dependent Variables y)

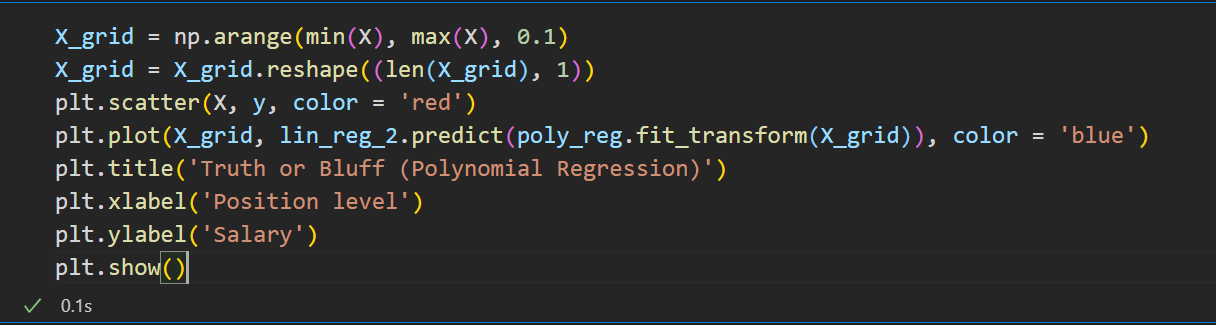


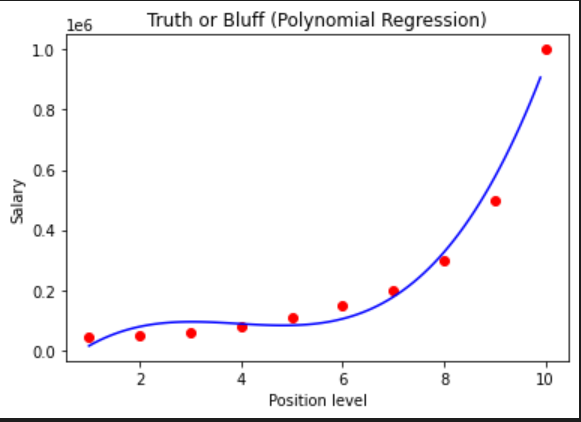
2.3 ຈົ່ງປ່ຽນຕົວປ່ຽນເອກະລາດໃຫ້ເປັນ Polynomial Features ໃຫ້ degree = 3 ແລະທຳການ (fit) ໂມເດວ Linear Regression ດ້ວຍ .



2.4 ຈົ່ງສະແດງຂໍ້ມູນ (X, y) ດ້ວຍ scatter Graph ບົນພື້ນຖານຊຸດຄຳສັ່ງ matplotlib

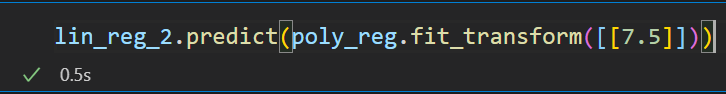
# ການສະແດງຜົນຂອງ Polynomial Regression (ສໍາລັບຄວາມລະອຽດທີ່ສູງຂຶ້ນແລະເສັ້ນໂຄ້ງທີ່ລຽບກວ່າ)





2.5 ຈົ່ງພະຍາກອນຄາດເດົາເງິນເດືອນ (Salary) ຂອງພະນກງານທີ່ມີຕໍາແໜ່ງລະດັບ (Level) 7.5?

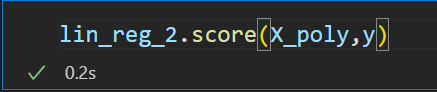
## ຄາດຄະເນຜົນໄດ້ຮັບໃຫມ່ດ້ວຍການ Polynomial Regression



array([240915.35547786])

2.6 ຈົ່ງຄິດໄລ່ R squared (ຄວາມຖືກຕ້ອງໃນການຄາດເດົາ)

# Polynomial Regression



0.9812097727913367