

Nama	: Adhimas Aryo Bimo
NIM	: 13523052
Model	: Polynomial Regressor

1. Sesuai dengan namanya ya Polynomial Regression. Model ini akan membuat fitur polinomial berderajat N dari sebuah dataset untuk menentukan prediksi. Dalam hal ini, algoritma Polynomial Regression yang saya buat memiliki 2 fungsi optimasi, yakni Gradient Descent dan Newton Method.

Tahap-tahapan dari algoritma tersebut adalah sebagai berikut:

a. Transformasi fitur

- Input asli dari X akan diubah menjadi fitur polinomial hingga berderajat tertentu. Dan tiap fitur pada data dipakai dengan pemilihan secara kombinasi

b. Training

- Data X yang telah diubah ke bentuk polinomial, dipilih metode optimasinya, untuk hal tersebut dapat digunakan Gradient Descent dan Newton Method
- Gradient Descent
 - o Hasil prediksi dari $y_{pred} = XW$ akan dihitung errornya lalu update bobot dengan rumus berikut :

$$W \leftarrow W - \eta \cdot \nabla L$$

- o L adalah MSE. Bisa ditambah dengan L1 atau L2
 - Newton Method
 - o Langsung update weight dengan rumus berikut:
- $$W = (X^T X + \lambda I)^{-1} (X^T y)$$
- Sederhananya, perbedaan antara keduanya hanya pada perhitungan weightnya saja. GD menghitung loss melalui turunan pertama dari error sedangkan pada NM menggunakan turunan ordo 2. Harapannya dengan menggunakan turunan ordo 2 ini akan lebih cepat dalam menemukan weight tanpa memerlukan tahapan berlebih. Namun ada konsekuensinya juga, NM secara komputasi perhitungannya lambat dibanding GD, Cuma GD butuh step lebih banyak

c. Prediksi

- Input baru ditransformasi jadi fitur polinomial berdasarkan weight yang telah dihitung selama train

$$y_{pred} = X_{poly} W$$

2. Ada perbedaan yang menarik antara model yang saya buat dengan model buatan scikit-learn. Pada hal ini, scikit-learn tidak ada model asli untuk polynomial, namun bisa dibuat pipe dari PolynomialFeatures dengan LinearRegressor. Selain

itu secara default, optimisasi pada scikit-learn melakukan OLS (Ordinary Least Square) dengan metode SVD sedangkan saya optimisasinya dengan GD. OLS ini akan menemukan nilai error pada training secara quote unquote “tepat” menggunakan newton method. Bisa jadi terjadi overfit. Dan ternyata ketika parameter model saya diganti pakai newton method hasilnya mirip.

Default:

===== PolyReg_custom =====	===== PolyReg_sklearn =====
Fold 1: 984.880	Fold 1: 990.413
Fold 2: 1139.776	Fold 2: 1350.856
Fold 3: 1075.629	Fold 3: 1107.173
Fold 4: 1022.872	Fold 4: 1045.111
Fold 5: 1113.245 Mean RMSE:	Fold 5: 1153.493 Mean RMSE:
1067.280 ± 56.921	1129.409 ± 123.725

Ubah parameter optimization pada PolynomialRegressor punya saya dengan newton method:

===== PolyReg_custom =====	===== PolyReg_sklearn =====
Fold 1: 990.438	Fold 1: 990.413
Fold 2: 1350.899	Fold 2: 1350.856
Fold 3: 1107.172	Fold 3: 1107.173
Fold 4: 1044.959	Fold 4: 1045.111
Fold 5: 1153.638	Fold 5: 1153.493 Mean RMSE:
Mean RMSE: 1129.421 ± 123.761	1129.409 ± 123.725

3. Saran saya untuk meningkatkan performa bisa diperbanyak degree nya untuk melihat pola kompleks yang ada di data.