

Program Fresh Graduate Academy Digital Talent Scholarship 2019 | Machine Learning

# Non-Linear Regression

M. Ramli & M. Soleh





Bagian 1

# Pendahuluan

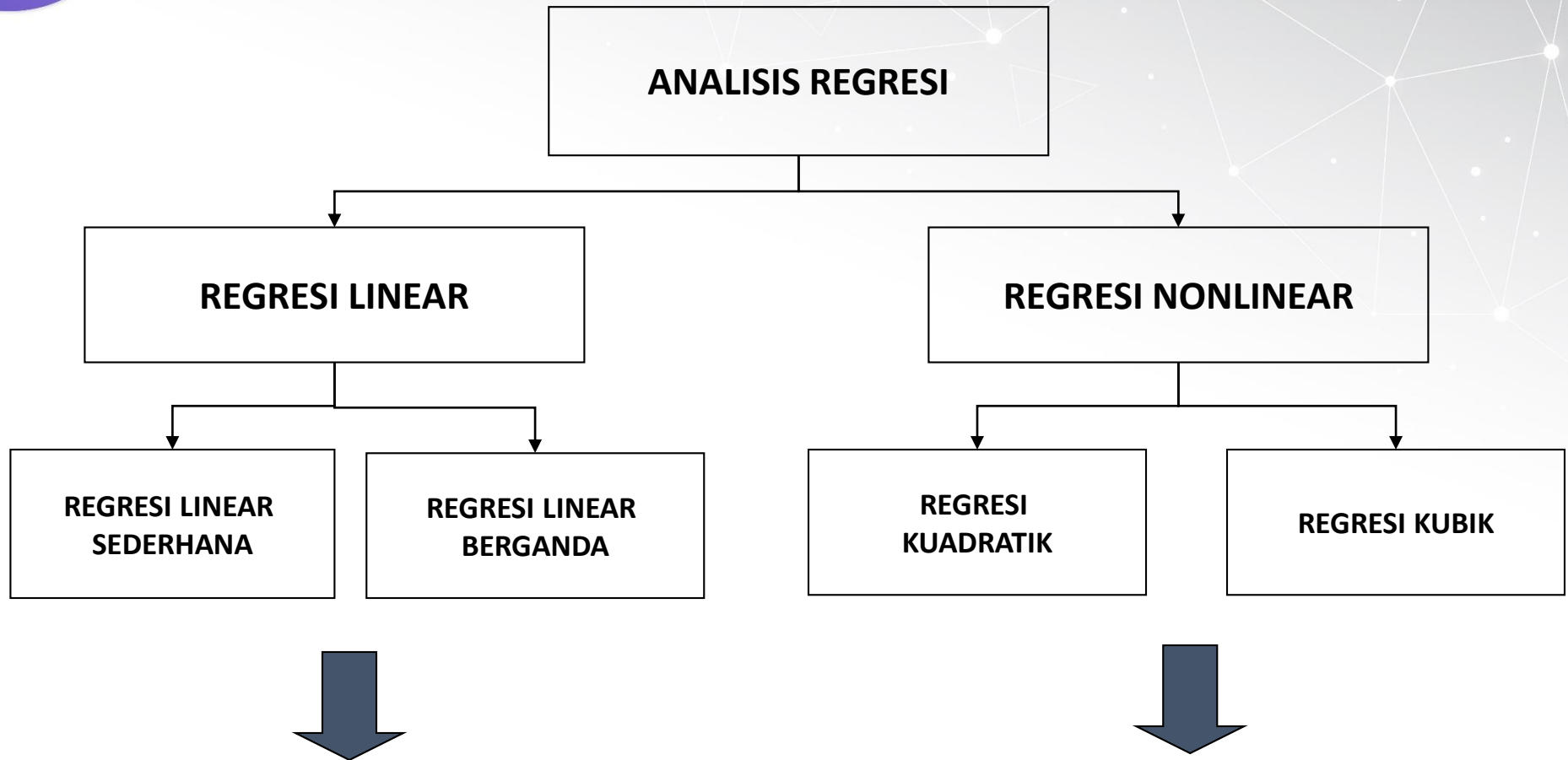


# Pendahuluan

- **Regresi non linier** adalah suatu metode untuk mendapatkan model non linier yang menyatakan hubungan variabel dependen dan variabel independen
- Regresi nonlinier dapat mengestimasi model hubungan variabel dependen dan independen dalam bentuk non linier dengan keakuratan yang lebih baik daripada regresi linier, karena dalam mengestimasi model dipakai iterasi algoritma.



# Pendahuluan (2)



**Membentuk Garis Lurus**

**Membentuk Garis Lengkung**

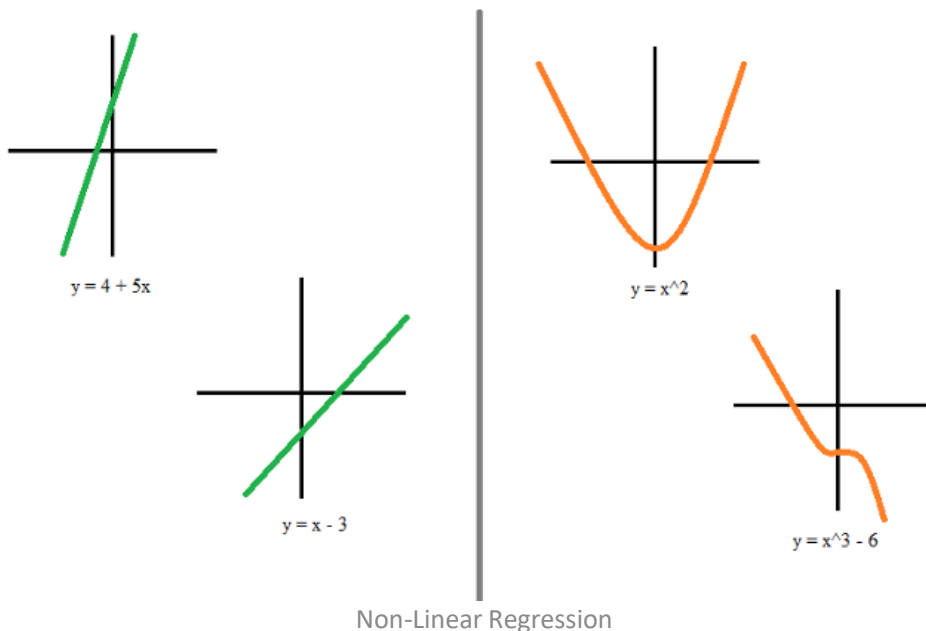


Bagian 2

# Model Regresi Non-Linear

# Linear VS Non-Linear

- Bagaimana Anda bisa menyatakan bahwa dataset yang ada, secara alamiah bersifat linear atau non-linear?
- Model yang dipilih untuk regresi, sangat bergantung kepada dataset itu sendiri.
- Mari tinjau kembali perbedaan fungsi linear dan non linear:



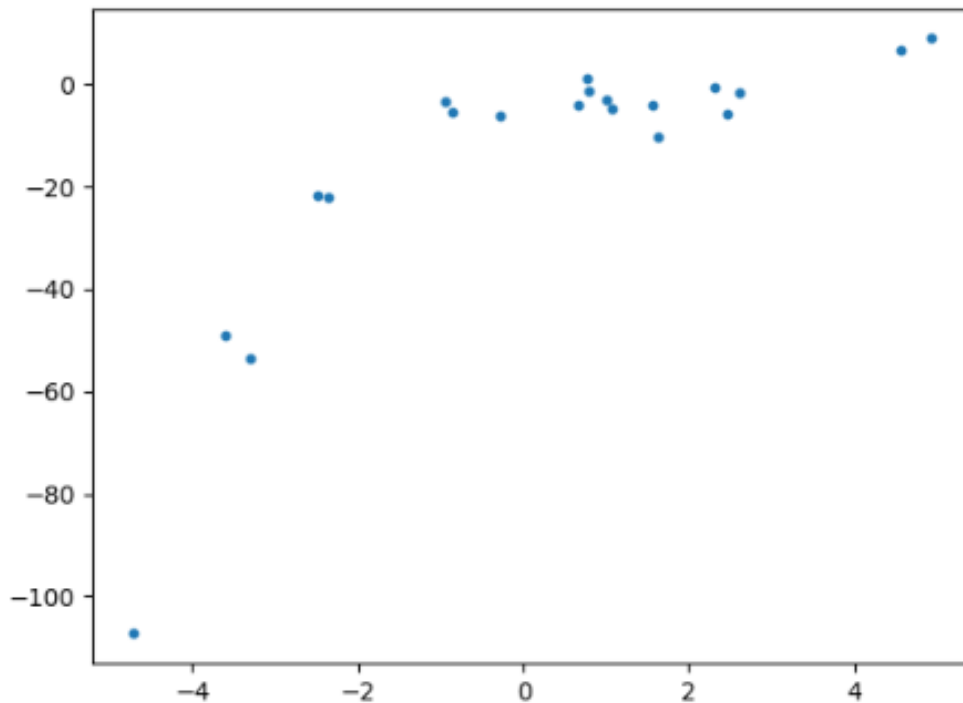


# Linear VS Non-Linear (2)

- **Fungsi linear:** Secara sederhana didefinisikan sebagai fungsi yang mengikuti prinsip:
  - Input/output = konstan
- Persamaan linear selalu merupakan polinomial derajat 1 (misalnya  $x + 2y + 3 = 0$ ).
- Dalam kasus dua dimensi, fungsi linear selalu membentuk garis
- Pada dimensi lain, fungsi linear mungkin juga membentuk bidang atau titik. "Bentuk" fungsi linear selalu benar-benar lurus, tanpa kurva apa pun.

# Mengapa Regresi Non-Linear

- Misalkan diberikan data seperti ini:

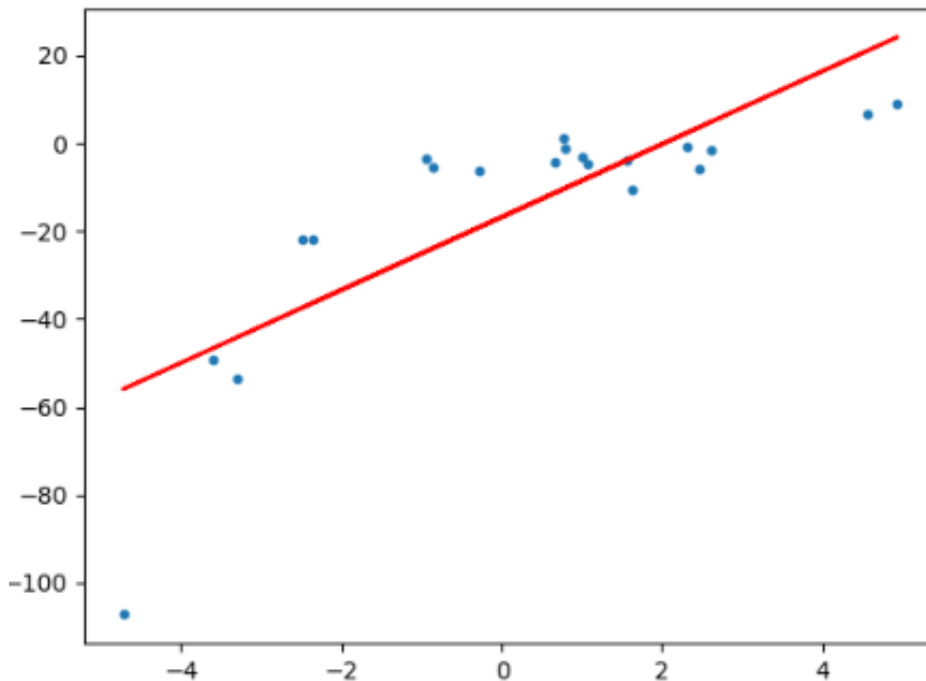


- Gunakan regresi linear atau non-linear?



# Mengapa Regresi Non-Linear (2)

- Digunakan model linear/ derajat 1:



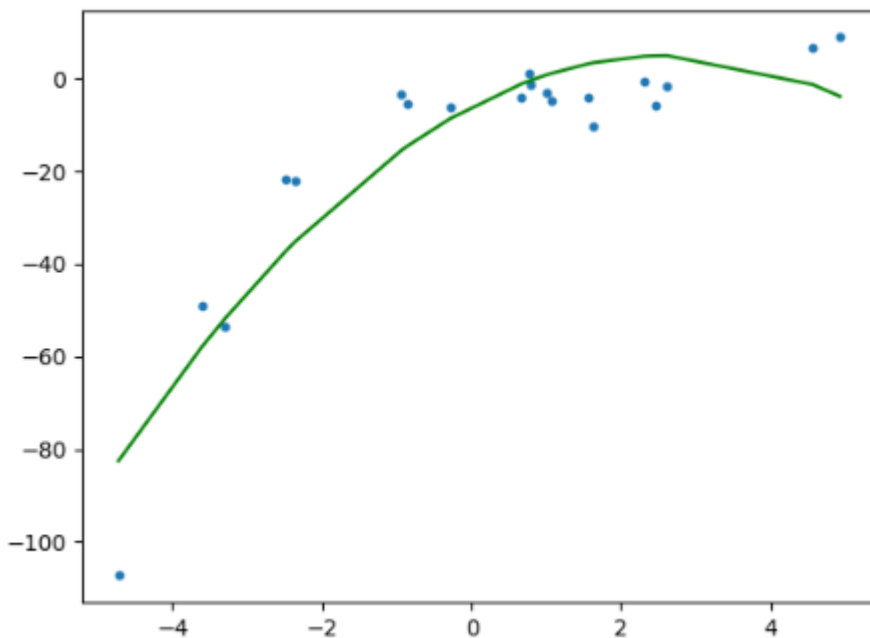
RMSE of linear regression  
is 15.908242501429998.

R2 score of linear  
regression is  
**0.6386750054827146**

- Banyak data yang “jauh” dari garis model.
- Kondisi: **under-fitting**

# Mengapa Regresi Non-Linear (3)

- Jika menggunakan model non-linear (kuadratik) / derajat 2:



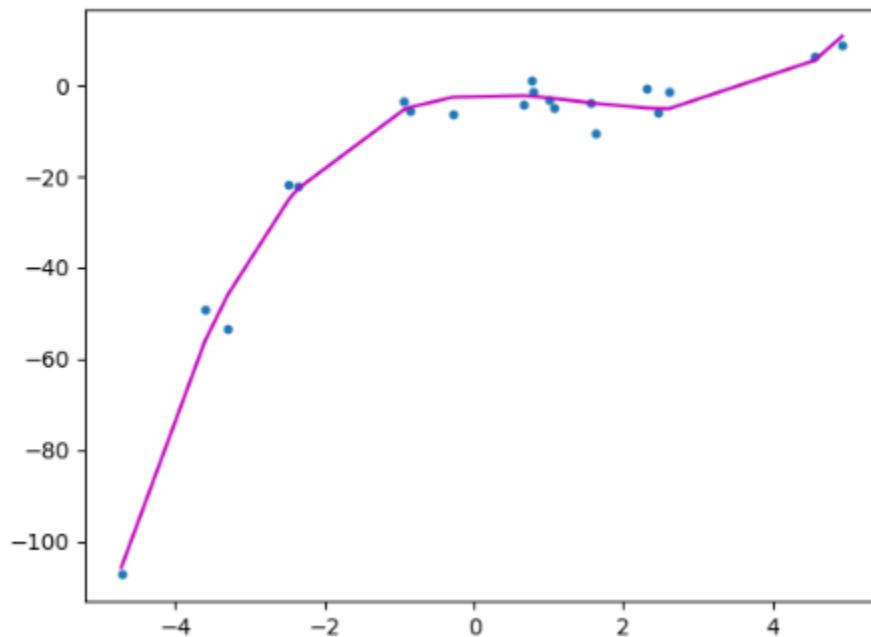
RMSE of polynomial  
regression is  
**10.120437473614711.**

R2 of polynomial  
regression is  
**0.8537647164420812.**

- Garis model terlihat sudah mendekati persebaran data.

# Mengapa Regresi Non-Linear (4)

- Jika menggunakan model non-linear (kubik) / derajat 3:



RMSE is

3.449895507408725

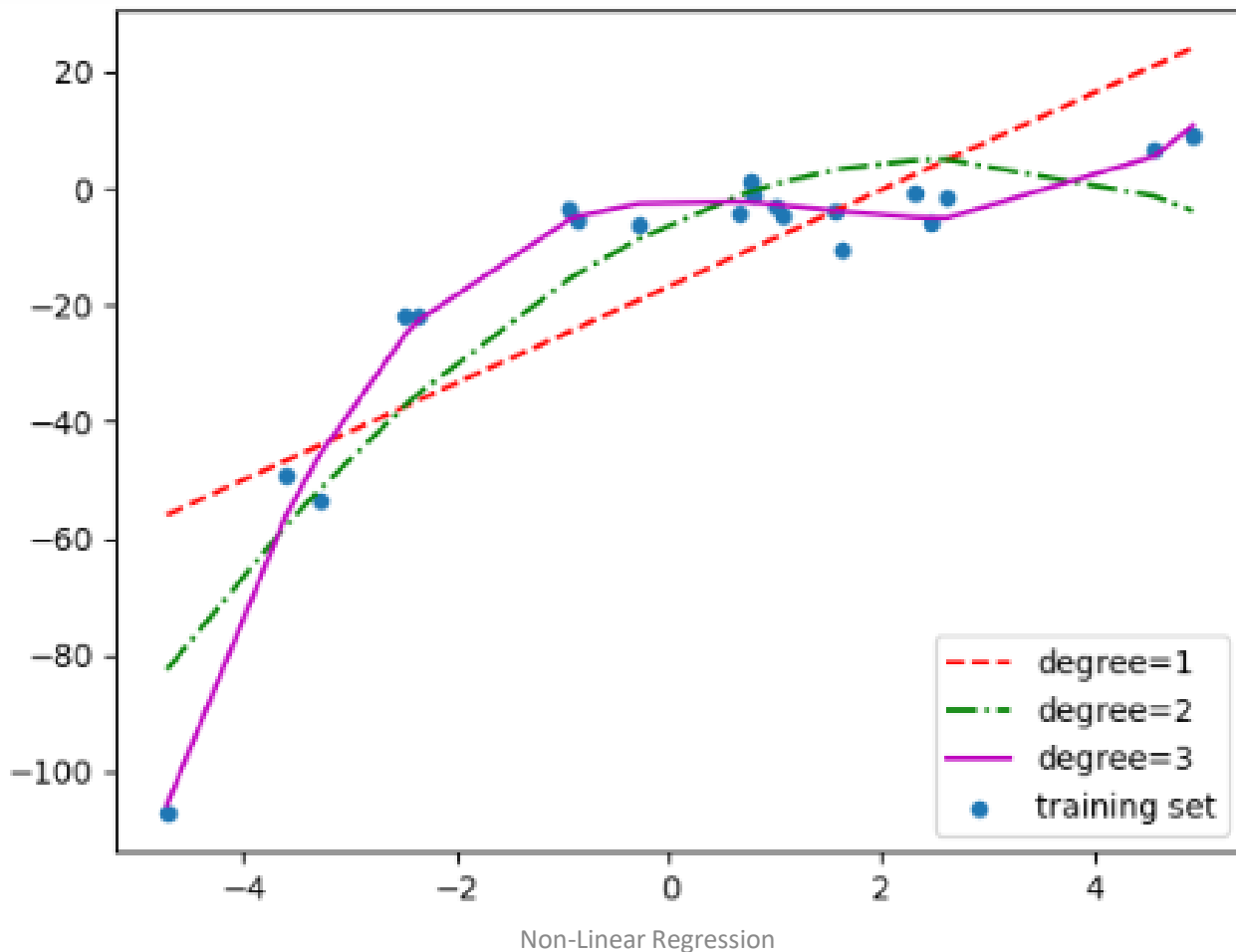
R2 score is

**0.9830071790386679**

- Garis model terlihat semakin mendekati persebaran data.
- Kondisi: **appropriate-fitting / correct-fit**

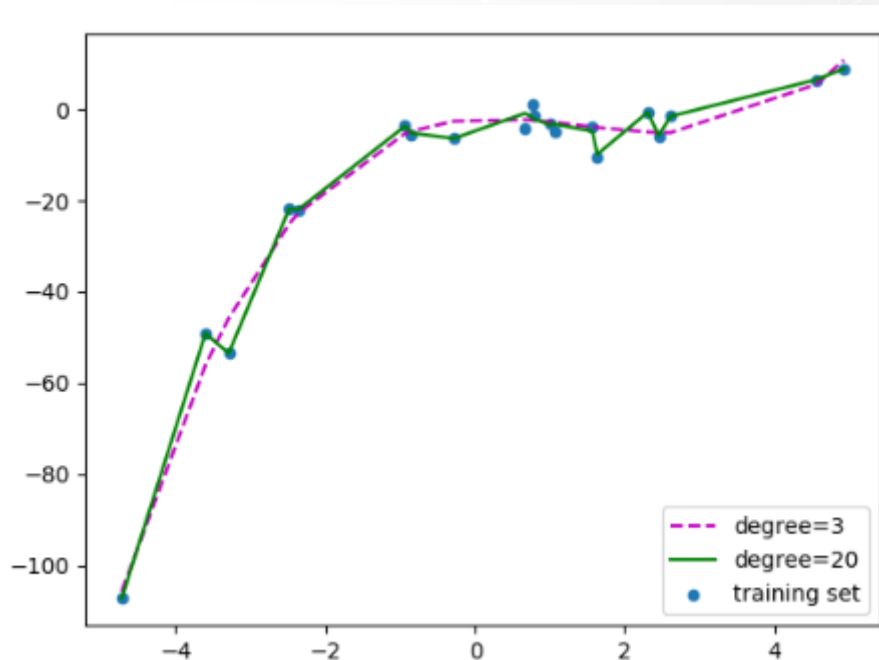
# Perbandingan 3 Model

- Berikut adalah perbandingan model linear, kuadratik dan kubik:



# Bagaimana Dengan Derajat Yang Lain

- Misalkan dimodelkan dengan derajat “20”:



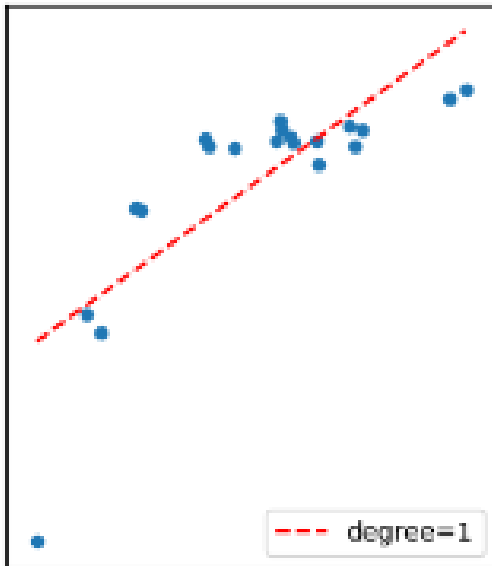
- Garis model juga melewati “noise” data.
- Kondisi: **over-fitting**



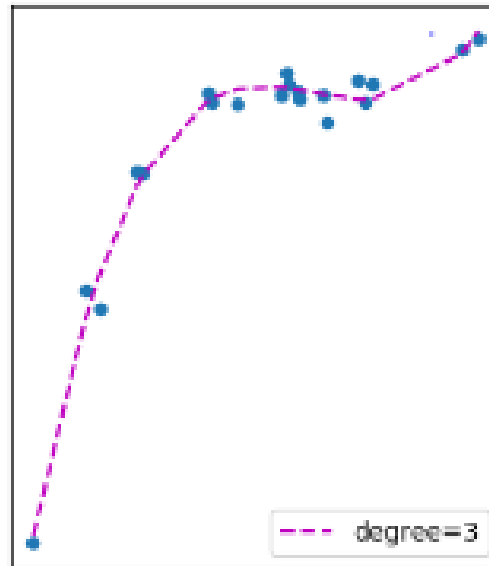
# Bias vs Varians

- **Bias merujuk pada kesalahan, karena asumsi model yang sederhana** dalam menyesuaikan data.
- **Bias yang tinggi** artinya bahwa model tidak dapat menangkap pola di dalam data dan ini mengakibatkan **kondisi under-fitting**.
- **Varians mengacu pada kesalahan, karena model yang kompleks** mencoba menyesuaikan data.
- **Varians tinggi** artinya model melewati sebagian besar titik pada data dan menghasilkan **kondisi over-fitting**

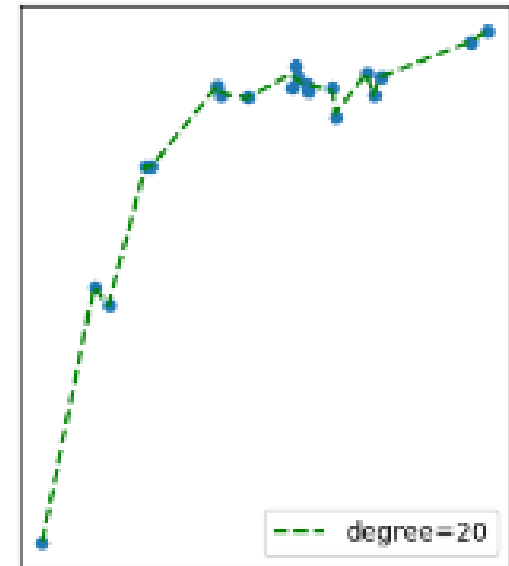
# Bias vs Varians (2)



Underfit  
High Bias  
Low Variance



Correct Fit  
Low Bias  
Low Variance



Overfit  
Low Bias  
High Variance



Bagian 2

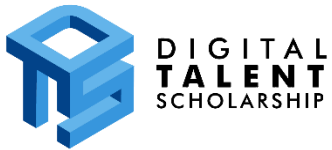
# Praktikum Lab




ML0101EN-Reg-NoneLinearRegression-py-v1.ipynb





IKUTI KAMI



-  [digitalent.kominfo](https://www.facebook.com/digitalent.kominfo)
-  [digitalent.kominfo](https://www.instagram.com/digitalent.kominfo)
-  [DTS\\_kominfo](https://twitter.com/DTS_kominfo)
-  Digital Talent Scholarship 2019

Pusat Pengembangan Profesi dan Sertifikasi  
Badan Penelitian dan Pengembangan SDM  
Kementerian Komunikasi dan Informatika  
Jl. Medan Merdeka Barat No. 9  
(Gd. Belakang Lt. 4 - 5)  
Jakarta Pusat, 10110

