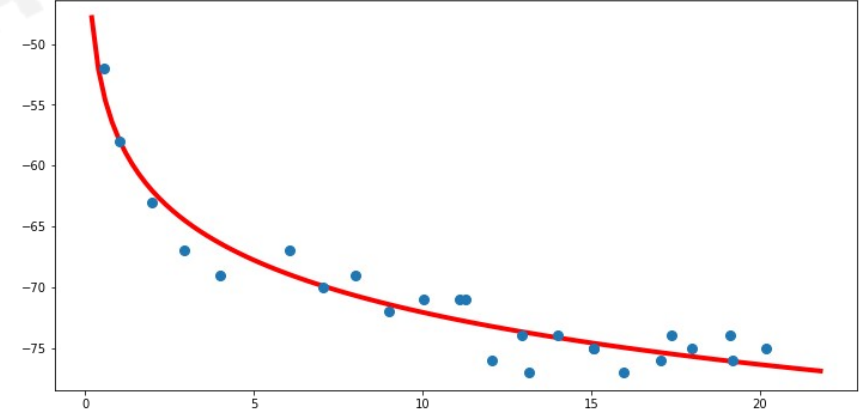
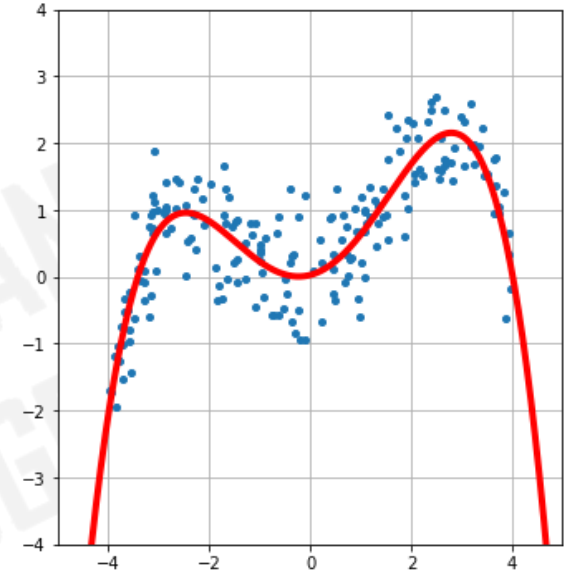
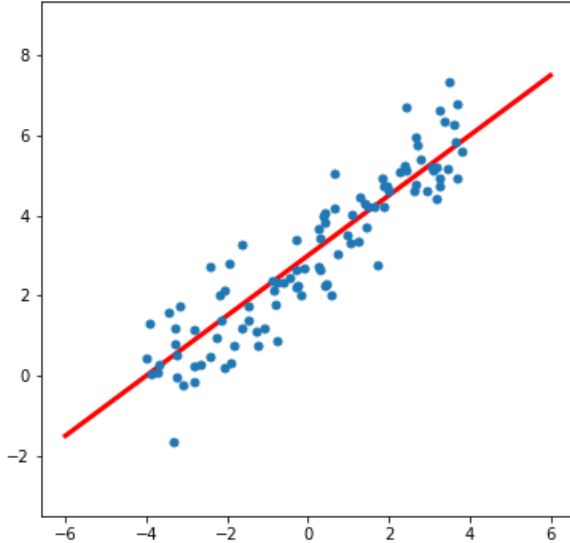


# Regresyon

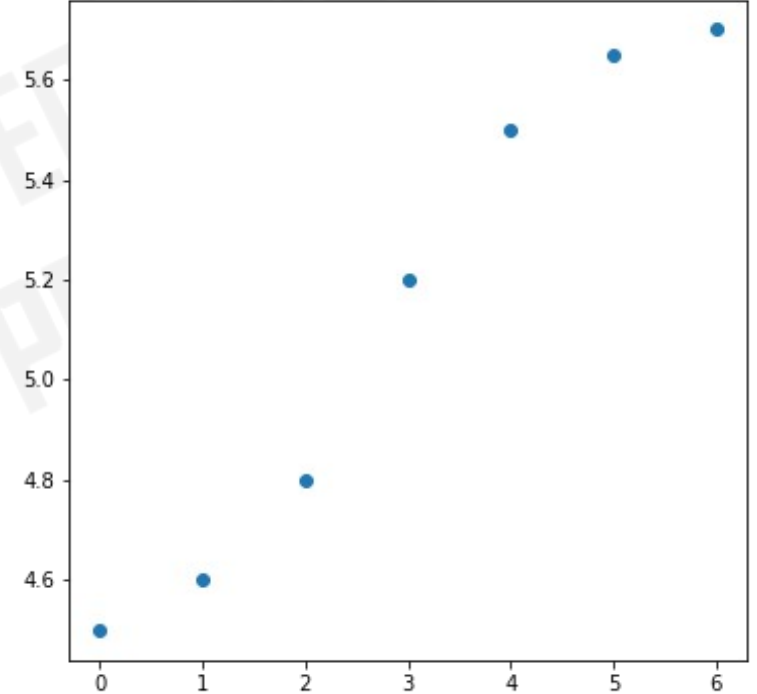
## Bölüm II

Serhan Daniş



# Polinom (Eğri) Regresyonu

- Grafikteki verileri inceleyelim.
- Bu veriler üzerine doğru mu yoksa eğri mi koymak daha uygun olacaktır?
- Eğri koyulacaksa nasıl bir eğri olmalı?
- Nasıl karar verilir?



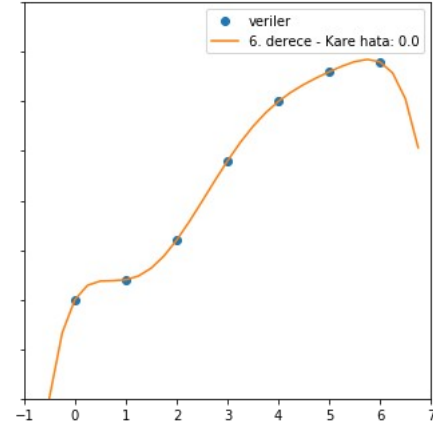
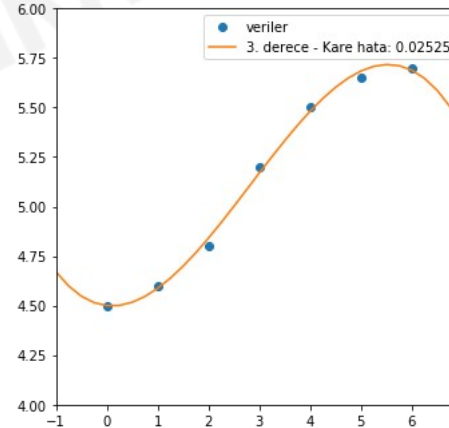
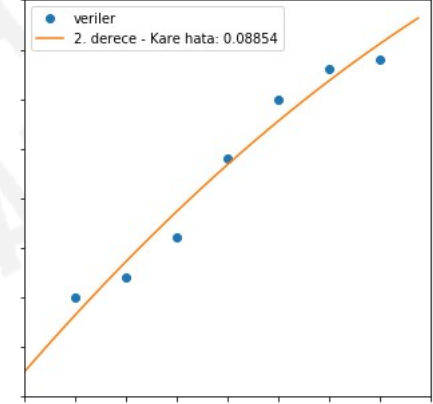
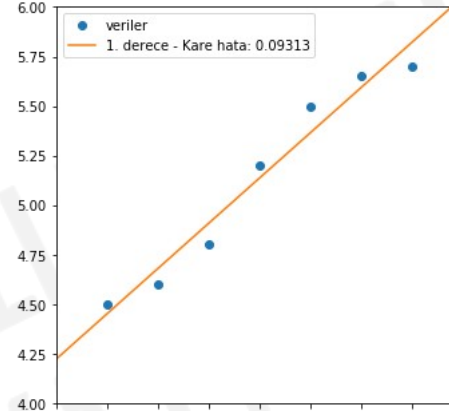
# Polinom

$$y = b_0 x^0 + b_1 x^1 + b_2 x^2 + b_3 x^3 + \dots = \sum_{i=0}^N b_i x^i$$

- Doğru denklemi daha genel bir halde yazılabilir.
- Her yeni terim eğriye bir büküm daha sağlayacaktır.
- Eğitim (yani katsayıları bulmak) için hala en küçük kareler yöntemi kullanılabilir.

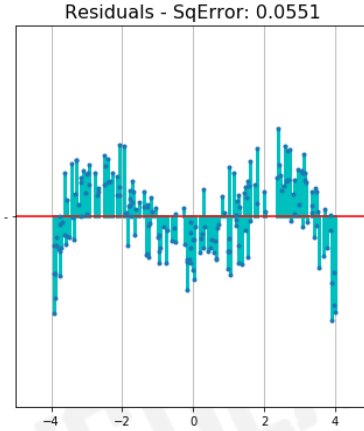
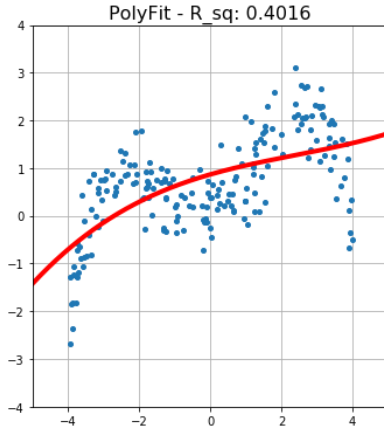
# Polinom derecesi

- **Polinomun derecesini** ne kadar artırırsak **hatayı** o kadar **düşürürüz**.
- **Veri sayısı kadar** artırabilirsek, hata **sıfıra** yaklaşacaktır.
- Ancak, bu duruma verileri ezberlemek veya aşırı öğrenme, **overlearning** veya **overfitting** deniyor.
- Amaç hatayı düşürmekten çok mümkün olan **en genel model parametresini**, bu durumda en uygun dereceyi bulmak.
- Henüz **görülmemiş veriler** için de hatanın düşük olmasını sağlamak gerekiyor.

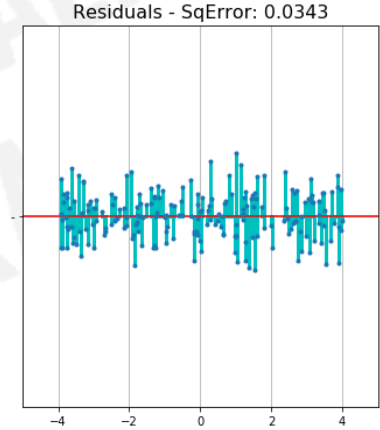
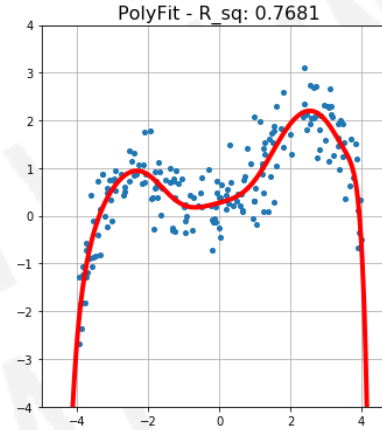


# Hata Analizi

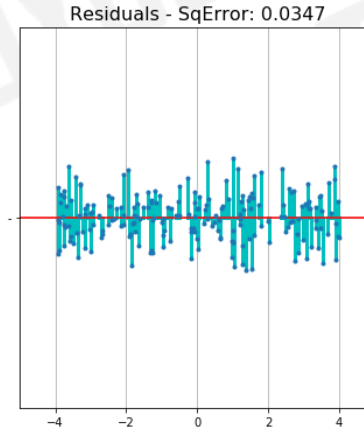
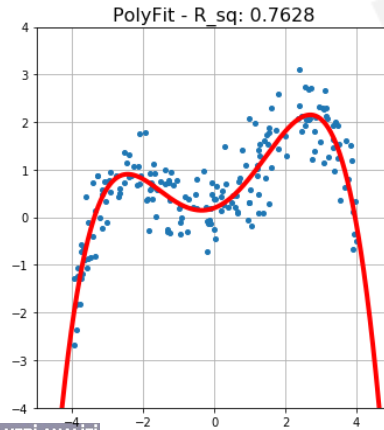
Degree: 3



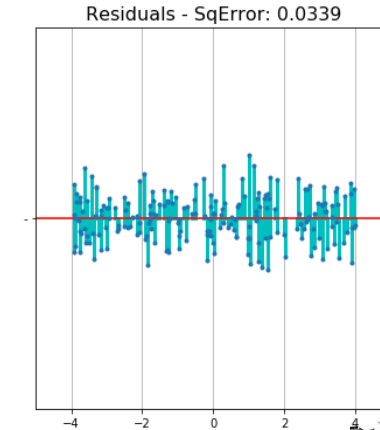
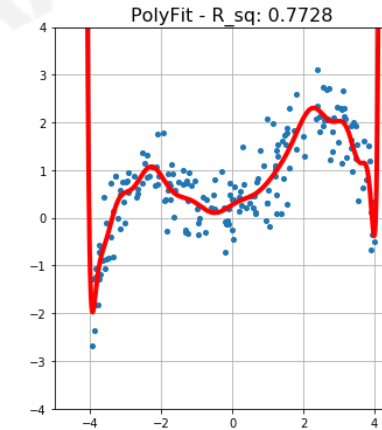
Degree: 14



Degree: 6



Degree: 20



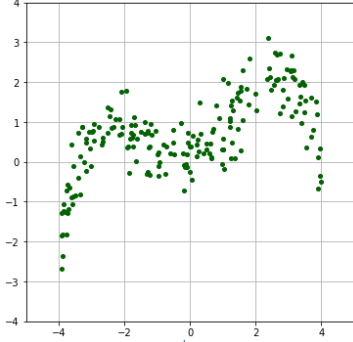
# Model Doğrulama

- **Denetimli öğrenmenin** (supervised learning) en çok başvurduğu yöntemdir.
- Aranan genelleştirilmiş alt model parametrelerinin **eğitimde kullanılmamış** veriler sayesinde bulunması sağlanır.
- Veriler **homojenliği** koruyacak şekilde **iki kümeye** ayrılır:
  - Eğitim (**training**) kümesi
  - Doğrulama (**validation**) kümesi
- **Eğitim kümesi** ile verilere göre **model parametreleri** aranır.
- **Doğrulama kümesindeki hatayı en düşük** yapan model parametreleri seçilir.

Veriler:

```
[[-0.748 0.207]
 [ 2.833 2.804]
 [ 0.227 0.337]
 [-0.318 -0.089]
 [ 1.463 1.121]
 [-2.074 0.764]
 [-1.592 0.409]
 [ 0.979 0.452]
 [-1.156 0.498]
 [-1.106 0.321]
 [ 0.094 -0.569]
 [-2.67 1.037]
 [-3.895 -1.292]
 [-2.386 1.418]
 [ 0.891 -0.247]
 [ 1.074 0.932]
 [ 1.488 1.424]
 [-1.403 0.655]
 [-1.123 0.827]
 [-3.793 -1.011]
 [-1.174 -0.216]
 [ 1.218 0.994]
 [-3.952 -2.271]
 [-0.473 0.28 ]
 [-3.729 -0.785]
 [ 0.822 0.537]
 [-1.629 0.411]
 [-0.127 0.213]
 [-3.171 0.908]
 [-1.551 0.601]
 [ 1.063 0.446]
```

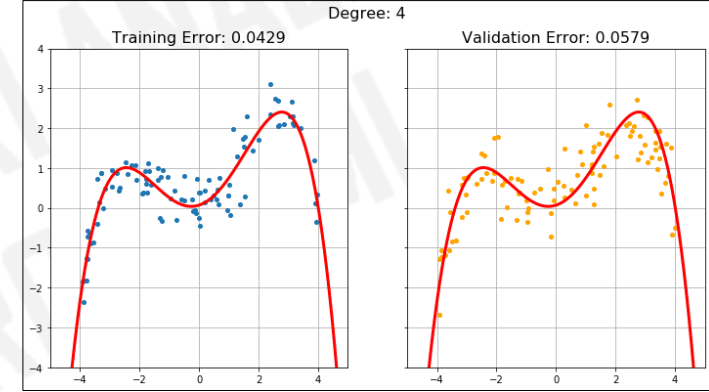
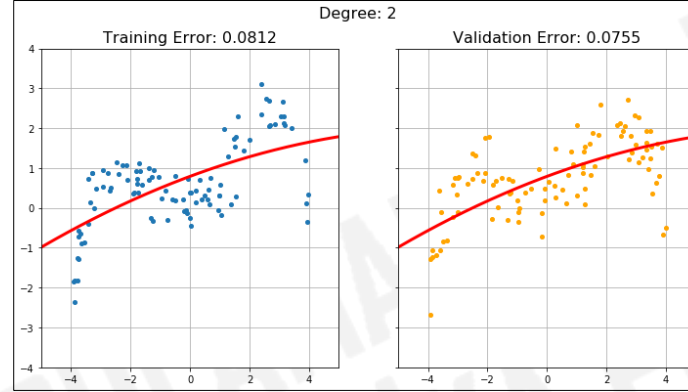
Verileri iki homojen kümeye bölüyoruz.



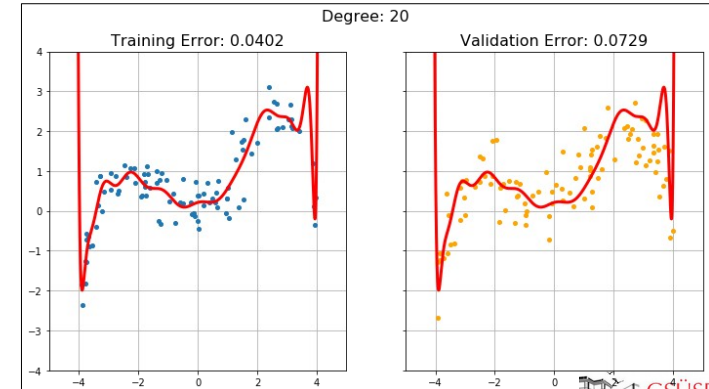
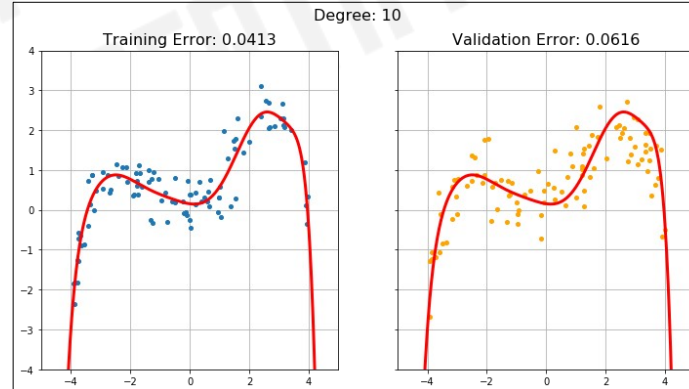
```
[ -3.246 0.72 ]
 [ 3.51 1.721]
 [ 0.837 0.673]
 [-0.784 0.054]
 [ 3.285 1.606]
 [-3.987 -1.881]
 [-2.768 0.925]
 [ 3.281 2.367]
 [ 2.291 1.972]
 [-2.704 1.144]
 [-1.763 0.671]
 [ 3.525 0.883]
 [ 1.449 0.896]
 [ 2.867 1.856]
 [ 1.001 0.846]
 [-0.685 -0.131]
 [ 0.167 0.302]
 [-3.82 -1.635]
 [ 0.084 0.349]
 [-1.747 1.06 ]
 [ 0.138 0.188]
```



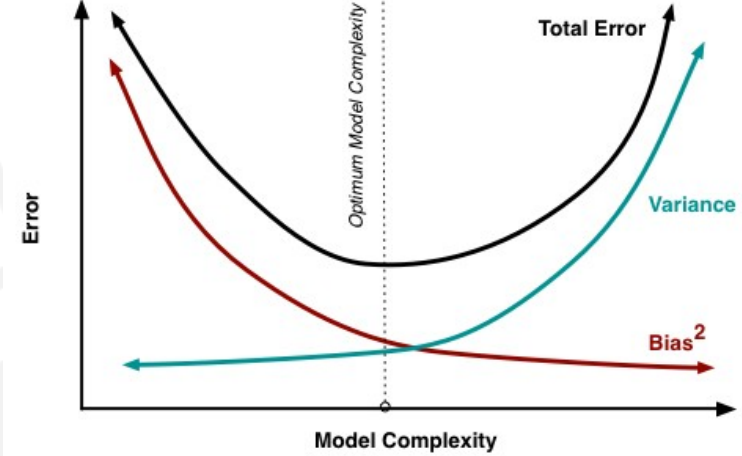
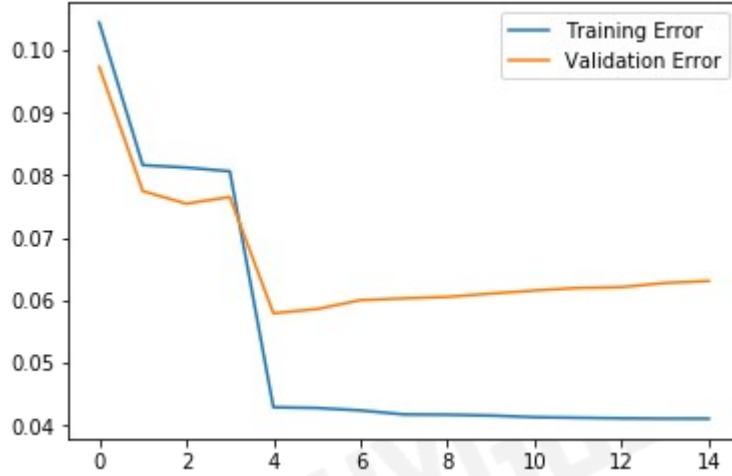
# Model Doğrulama



Doğrulama hatasını (Validation Error) en küçük veren model parametresini (örn. power or degree) seçiyoruz.



# Model Doğrulama



- Amaç, veri kümesinin en genel halinde de hatayı en düşük yapacak model parametrelerini (hyper parameters) bulmak.
- Genel yönetime göre, model karmaşıktıkça eğitim verileri (bias = yanlılık) hatası devamlı olarak düşer, doğrulama verileri (variance = çeşitlilik) hatası önce düşer, bir noktadan sonra da artmaya başlar.
- Doğrulama hatasını bu en düşük noktaya getiren parametre (karmaşıklık derecesi) söz konusu verilere en uygun model parametresidir.



# Diğer regresyon çeşitleri

- En küçük kareler yöntemini kullanabilmek için katsayılar toplam durumunda olmalıdır. Bunu sağladığımız durumlarda yöntem her zaman kullanılabilir.

- Çok boyutlu doğrusal regresyon:

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \dots$$

- Çok boyutlu polinom regresyon:

$$y = b_0 + b_1 x_1 + b_{12} x_1^2 + b_2 x_2 + b_{22} x_2^2 + b_{21} x_1 x_2 + \dots$$

- Doğrusal olmayan (nonlinear) regresyon:

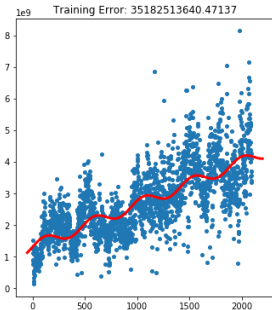
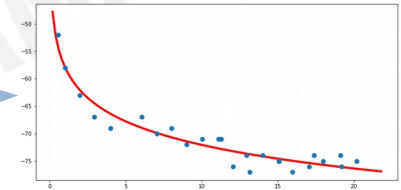
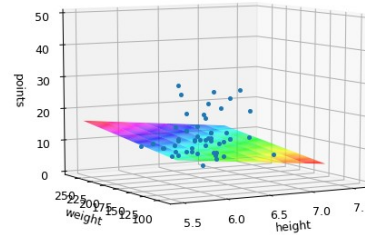
$$y = b_0 + b_1 \log(x) \quad \text{---} \quad y = b_0 + b_1 \sin(wx + z)$$

- Doğrusal olmayan (nonlinear) regresyon (?):

$$y = b_0 + b_1 \log(b_2 x)$$

$$y = b_0 + b_1 \sin(b_2 x)$$

$$y = b_0 + \frac{b_1 x_1}{b_2 x_2 + b_3 x_3}$$



# Örnekler ve Çalışma

