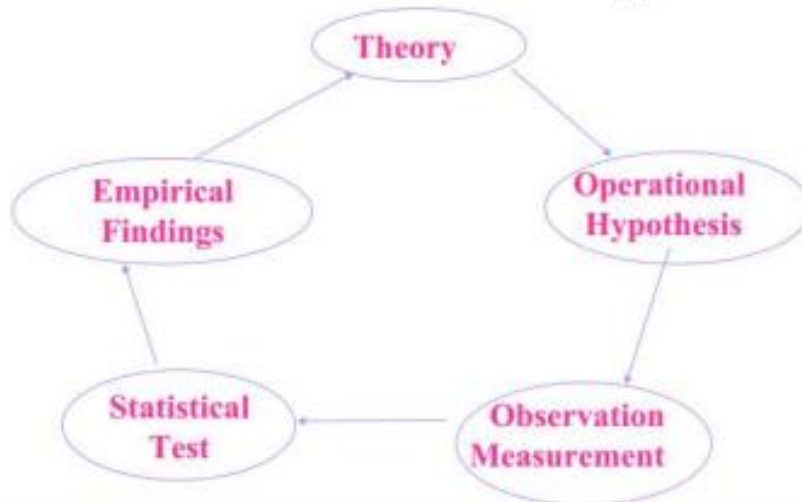


# Data Modelling and Regression Techniques

M. Fatih Amasyalı

## The traditional scientific approach





## ChatGPT

Bu görsel, "Data Modelling and Regression Techniques" (Veri Modelleme ve Regresyon Teknikleri) başlıklı bir sunumun veya ders notlarının iki slaytını içeriyor. İlk slaytta, belgenin başlığı ve "M. Fatih Amasyalı" ismi yer alıyor. Muhtemelen bu kişi, dersi veren öğretim üyesi veya notları hazırlayan kişidir.

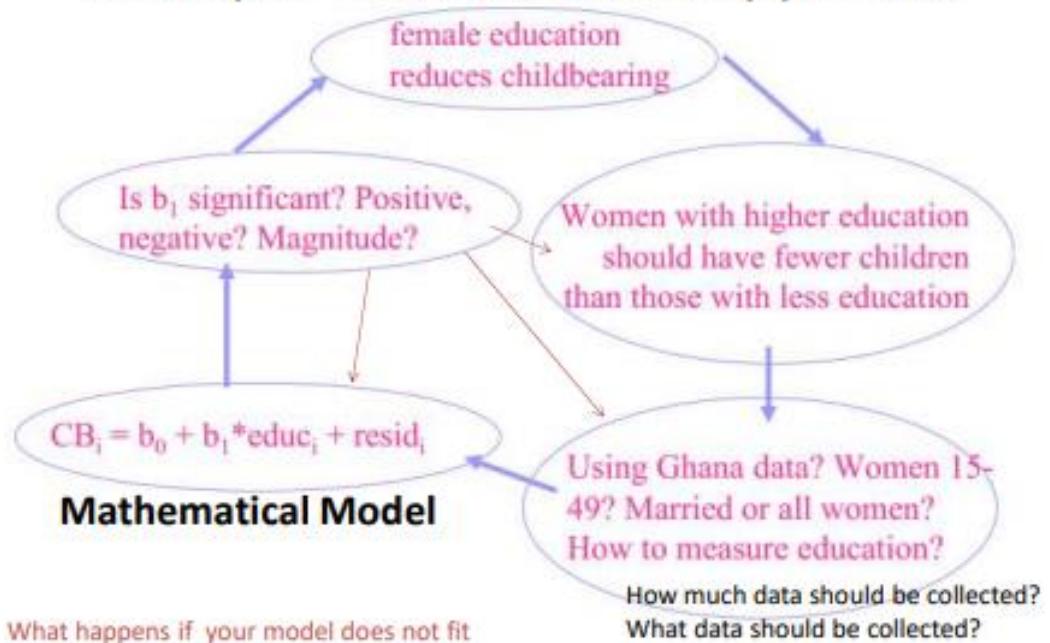
İkinci slaytta "The traditional scientific approach" (Geleneksel bilimsel yaklaşım) başlığı altında bir konsept şeması bulunuyor. Şema, bilimsel sürecin dört ana adımını gösteriyor:

1. **Theory (Teori):** Bilimsel bir teori veya varsayım.
2. **Empirical Findings (Ampirik Bulgular):** Gözlemler ve deneyler sonucunda elde edilen veriler.
3. **Statistical Test (İstatistiksel Test):** Elde edilen verilerin istatistiksel yöntemlerle analiz edilmesi.
4. **Operational Hypothesis (Operasyonel Hipotez):** Test edilebilir ve ölçülebilir bir hipotez.

Şemada gösterilen bu döngü, bilimsel araştırmanın nasıl teoriden empirik bulgulara, oradan istatistiksel testlere ve sonunda operasyonel hipotezlere ilerlediğini temsil ediyor. Bu döngü, bilimsel bilginin nasıl sistemli bir şekilde geliştirildiğini ve doğrulandığını gösteriyor.



## Example of a scientific approach



Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG DEPT.

- A model is an underlying theory about how the world works. It includes:
  - Assumptions
  - Key players (independent variables)
  - Interactions between variables
  - Outcome set (dependent variables)
- $CB = x_1 + educ * x_2 + resid$ 
  - Assumptions?, variables?, interactions?

Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG DEPT.



## ChatGPT

Bu görsel, bilimsel bir yaklaşımın örneğini ve bir matematiksel modeli anlatan bir ders notu veya sunum slaytıdır. Görselde iki ana bölüm bulunuyor:

### Üst Bölüm:

Burada, kadınların eğitim seviyesi ile çocuk sahibi olma sayısı arasındaki ilişkiyi inceleyen bir bilimsel soruşturma örneği gösteriliyor. "female education reduces childbearing" (kadın eğitimi çocuk sahibi olmayı azaltır) varsayımından hareketle, daha yüksek eğitilmiş kadınların daha az çocuğa sahip olması gerektiği üzerine bir hipotez kurulmuş. Ardından bu hipotezi test etmek için bir matematiksel model öneriliyor: " $CB = b_0 + b_1 \cdot educ + resid$ ", burada 'CB' çocuk sahibi olma sayısını, 'educ' eğitim seviyesini ve 'resid' kalan varyansı (hata terimini) temsil ediyor. Modelin, Gana verileri kullanılarak ve belirli demografik gruplardan (örneğin 15-49 yaş arası, evli veya tüm kadınlar) toplanan verilerle nasıl test edileceği tartışılıyor.

### Alt Bölüm:

Bir modelin dünyanın nasıl işlediği hakkında altta yatan bir teori olduğu ve şunları içerdiği açıklanıyor:

- Varsayımlar
- Ana oyuncular (bağımsız değişkenler)
- Değişkenler arası etkileşimler
- Sonuç kümesi (bağımlı değişkenler)

Matematiksel modelin formülü tekrarlanmış ve bu modelde yer alması gereken varsayımlar, değişkenler ve etkileşimlerin neler olması gerektiği soru olarak sunulmuş.

Slaytın altındaki metin kısmında, modelin içermesi gereken unsurlar detaylandırılmış ve modelin yapısını oluşturan matematiksel ifadenin ne olduğu üzerinde durulmuş. Modelin varsayımlarını, değişkenlerini ve etkileşimlerini sorgulayan sorularla kapanış yapılmış. Bu, bir modelin geçerliliğini ve güvenilirliğini değerlendirirken göz önünde bulundurulması gereken unsurları vurgulamaktadır.



## Regression Models

- Relationship between one dependent variable and explanatory variable(s)
- Use equation to set up relationship
  - Numerical Dependent (Response) Variable
  - 1 or More Numerical or Categorical Independent (Explanatory) Variables

## Regression Modeling Steps

1. Hypothesize Deterministic Component
  - Estimate Unknown Parameters
2. Evaluate the fitted Model
3. Use Model for Prediction & Estimation



## ChatGPT

Bu görselde regresyon modelleri ve regresyon modelleme adımları anlatılmaktadır.

### Regresyon Modelleri:

- Bağımlı (yanıt) değişken ile bir veya daha fazla bağımsız (açıklayıcı) değişken arasındaki ilişkiyi kurar.
- İlişkiyi kurmak için denklem kullanılır.
  - Sayısal Bağımlı (Yanıt) Değişken
  - Bir veya daha fazla Sayısal veya Kategorik Bağımsız (Açıklayıcı) Değişken

### Regresyon Modelleme Adımları:

1. **Deterministik Bileşenin Hipotezini Kurma:**
  - Bilinmeyen parametreleri tahmin etme.
2. **Uygun Modeli Değerlendirme:**
  - Modelin uygunluğunu değerlendirme.
3. **Modeli Tahmin ve Kestirim İçin Kullanma:**
  - Modeli veri tahmin etmek ve kestirim yapmak için kullanma.

Bu slayt, regresyon modelleme sürecinin temel kavramlarını ve adımlarını öğrencilere tanıtmak için kullanılan ders materyali gibi görünmektedir. Regresyon modelleri, değişkenler arasındaki ilişkileri anlamak ve gelecekteki değerleri tahmin etmek için kullanılır.

## Specifying the deterministic component

1. Define the dependent variable and independent variable(s)
2. Hypothesize Nature of Relationship
  - Expected Effects (i.e., Coefficients' Signs)
  - Functional Form (Linear or Non-Linear)
  - Interactions

## Model Specification Is Based on Theory

1. Previous Research
2. 'Common Sense'
3. Data (which variables, linear/non-linear etc.)





## ChatGPT

Bu görsel, regresyon analizinde deterministik bileşeni belirlemenin ve model spesifikasyonunun nasıl yapılacağını açıklayan bir ders notunu göstermektedir.

### Deterministik Bileşeni Belirtmek:

1. Bağımlı ve bağımsız değişken(ler)i tanımla.
2. İlişkinin doğasına dair hipotezler kur:
  - Beklenen etkiler (yani, katsayıların işaretleri)
  - Fonksiyonel form (Doğrusal veya doğrusal olmayan)
  - Etkileşimler

### Model Spesifikasyonu Teoriye Dayalıdır:

1. Önceki Araştırmalar
2. 'Sağduyu'
3. Veri (hangi değişkenlerin kullanılacağı, doğrusal/doğrusal olmayan ilişkiler vb.)

Bu slayt, bir modelin nasıl oluşturulacağını ve hangi teorik temellere dayanması gerektiğini anlatmaktadır. Özellikle, araştırmanın doğası hakkında sağlam bir hipotez geliştirmek ve bu hipotezi destekleyecek veri toplamak için kullanılacak metodolojiyi belirlemenin önemini vurgulamaktadır.

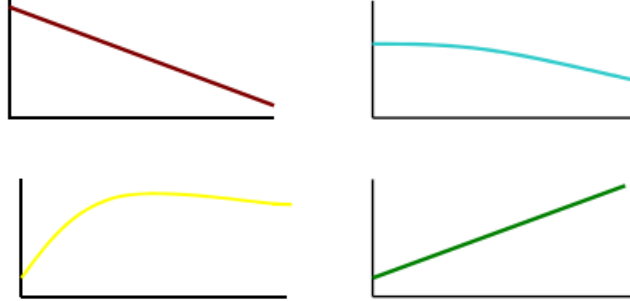




## Thinking Challenge: Which Is More Logical?

X: How many hours did you study in the night before the exam?

Y: your exam grade



## Types of Regression Models

The linear first order model  $Y = \beta_0 + \beta_1 X + \varepsilon$

The linear second order model  $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$

The linear n order model  $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_n X^n + \varepsilon$

$\varepsilon$  is random error.

The word **linear** refers to the linearity of the parameters  $\beta_i$ .

The **order** (or **degree**) of the model refers to the highest power of the predictor variable X.



## ChatGPT

Bu görselde bir düşünme meydan okuması ve regresyon modellerinin tipleri anlatılmaktadır.

### Düşünme Meydan Okuması:

Hangisi Daha Mantıklı?

- X: Sınavdan önceki gece kaç saat çalıştın?
- Y: Sınav notun

Burada dört farklı grafik gösterilmekte ve bu grafikler, sınavdan önce çalışma süresi (X) ile sınav notu (Y) arasındaki potansiyel ilişkileri temsil etmekte. Her grafik, çalışma süresinin sınav notuna etkisinin farklı olduğu bir senaryoyu tasvir etmekte: azalan, sabit, kavisli ve artan eğilimler.

### Regresyon Modellerinin Tipleri:

- Doğrusal birinci dereceden model:  $Y = \beta_0 + \beta_1X + \varepsilon$
- Doğrusal ikinci dereceden model:  $Y = \beta_0 + \beta_1X + \beta_2X^2 + \varepsilon$
- Doğrusal n dereceden model:  $Y = \beta_0 + \beta_1X + \beta_2X^2 + \dots + \beta_nX^n + \varepsilon$

$\varepsilon$ , rastgele hata terimidir. "Doğrusal" terimi,  $\beta_i$  parametrelerinin doğrusallığına işaret eder. Modelin "derecesi" veya "mertebesi", tahmin edici değişken X'in en yüksek kuvvetine göre belirlenir. Bu, regresyon modelinin karmaşıklığını ve hangi derecede polinom kullanıldığını gösterir.

## Types of Regression Models

- If the parameters are linear related to the each other the model is linear. A non-linear first order model:  $Y = (\beta_0 X) / (\beta_1 + X) + \epsilon$
- If  $X$  has  $d$  dimensions, a linear first order model:  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_d X_d + \epsilon$

- A linear first order model  $Y = \beta_0 + \beta_1 X + \epsilon$
- To get the model, we need to estimate the parameters  $\beta_0$  and  $\beta_1$
- Thus, the estimate of our model is

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

- $Y_{\text{hat}}$  denotes the predicted value of  $Y$  for some value of  $X$ , and  $\beta_{0\text{hat}}$  and  $\beta_{1\text{hat}}$  are the estimated parameters.



## ChatGPT

Bu görsel, regresyon modellerinin tiplerini anlatan bir ders notu bölümünü göstermektedir.

### Regresyon Modellerinin Tipleri:

- Eğer parametreler birbirleriyle doğrusal olarak ilişkiliyse model doğrusaldır. Doğrusal olmayan birinci dereceden bir model örneği:  $Y = (\beta_0 X) / (\beta_1 X) + \epsilon$
- Eğer  $X$  d boyutundaysa, doğrusal birinci dereceden bir model:  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_d X_d + \epsilon$

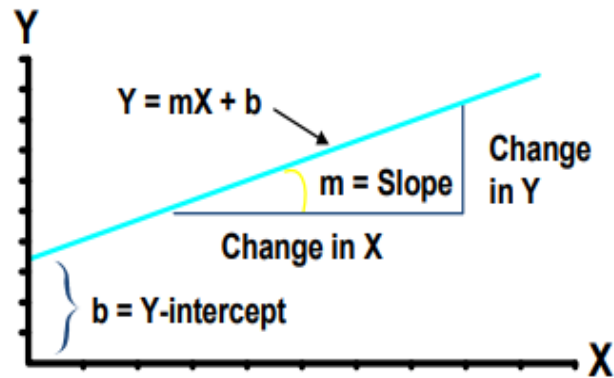
### Ayrıca:

- Doğrusal birinci dereceden bir model:  $Y = \beta_0 + \beta_1 X + \epsilon$
- Modeli elde etmek için,  $\beta_0$  ve  $\beta_1$  parametrelerini tahmin etmemiz gerekiyor.
- Bu yüzden modelimizin tahmini:  $\hat{Y} = \beta_0 + \beta_1 X$
- $\hat{Y}$ ,  $X$ 'in belirli bir değeri için  $Y$ 'nin tahmin edilen değerini ifade eder ve  $\beta_0$  hat ve  $\beta_1$  hat tahmin edilen parametrelerdir.

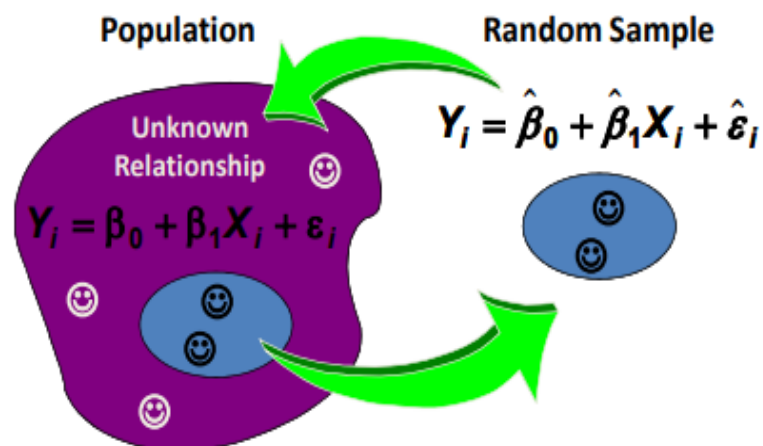
Bu notlar, bir regresyon modelinin nasıl kurulacağını ve parametrelerin nasıl tahmin edileceğini açıklar.  $Y$ 'nin  $X$ 'e bağlı değişikliğine modellemek için kullanılan matematiksel ifadeleri içerir.



## An Old Friend



## Population & Sample Regression Models



Mostly, you can not reach the all population.



## ChatGPT

Görselde, doğrusal regresyon analizi ve popülasyon ile örneklem regresyon modellerinin temel kavramları açıklanmaktadır.

### Birinci Bölüm: "Eski Bir Dost"

- Doğrusal denklemin temel formu gösterilmiş:  $Y = mX + b$
- Burada "m" X'in birim değişikliğine karşılık Y'de meydana gelen değişimi ifade eden eğimi (slope) temsil eder.
- "b" ise Y-ekseni kesim noktasını (Y-intercept) ifade eder.

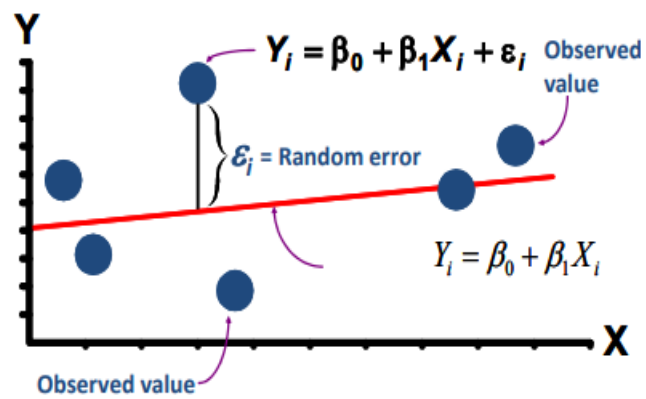
### İkinci Bölüm: "Popülasyon & Örneklem Regresyon Modelleri"

- Popülasyon içerisinde alınan rastgele bir örneklemin (Random Sample) nasıl popülasyonun genelini temsil ettiğini gösteren bir şema bulunmakta.
- Popülasyon için bilinmeyen ilişkiyi temsil eden formül:  $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$
- Rastgele örneklem için ilişkiyi temsil eden tahmini formül:  $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\epsilon}_i$
- Şemada, genellikle tüm popülasyona ulaşamayacağı için rastgele örneklemle çalışıldığı vurgulanıyor.

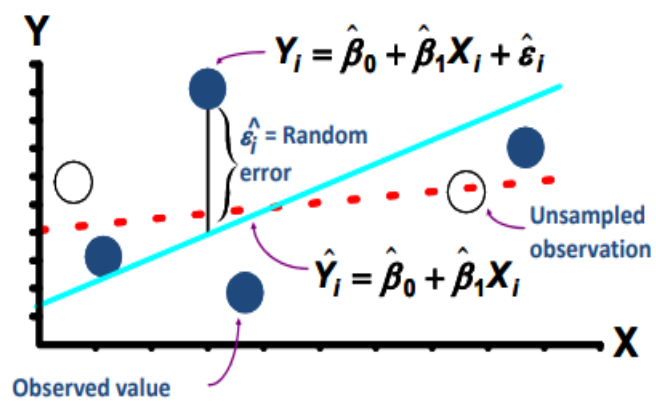
Bu görsel, regresyon analizinin temelini ve popülasyon ile örneklem arasındaki ilişkiyi anlamaya yönelik bir ders notu veya eğitim materyali parçası gibi görünmektedir.



## Population Linear Regression Model



## Sample Linear Regression Model







## ChatGPT

Bu görsel, popülasyon ve örneklem doğrusal regresyon modellerini gösteren iki grafik içermektedir.

### Üstteki Grafik - Popülasyon Doğrusal Regresyon Modeli:

- Bu model, popülasyonun tamamını temsil eden bir regresyon çizgisini gösterir.
- Gerçek veri noktaları (gözlemlenen değerler) ve regresyon çizgisi arasındaki dikey mesafeler, her bir gözlem için rastgele hata terimini ( $\epsilon_i$ ) temsil eder.
- Formül olarak,  $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$  şeklinde ifade edilir.
- $Y_i$ , her bir  $X_i$  değeri için bağımlı değişkenin gerçek değerini ifade eder.

### Altteki Grafik - Örneklem Doğrusal Regresyon Modeli:

- Bu model, seçilen bir örneklem üzerinde uygulanan regresyon çizgisini gösterir.
- Örneklemdeki gerçek veri noktaları ve regresyon çizgisi arasındaki dikey mesafeler, örneklemdeki her bir gözlem için rastgele hata terimini ( $\hat{\epsilon}_i$ ) temsil eder.
- Tahmini formül olarak,  $\hat{Y}_i = \beta_0 + \beta_1 X_i + \hat{\epsilon}_i$  şeklinde ifade edilir.
- $\hat{Y}_i$ , model tarafından tahmin edilen değerleri ifade ederken, kırmızı noktalı çizgi örneklenmemiş gözlemleri temsil eder.

Her iki grafik de doğrusal regresyon analizinin temel kavramlarını görselleştirir: gerçek veri noktaları, tahmin edilen regresyon çizgisi ve bu çizgiye göre rastgele hatalar.



## Estimating Parameters: Least Squares Method

## Least Squares

- 1. 'Best Fit' means difference between actual Y values & predicted Y values are a minimum. *But* positive differences off-set negative. So square errors!

$$\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n \hat{\varepsilon}_i^2$$

- 2. LS Minimizes the Sum of the Squared Differences (errors) (SSE)

- Mean squared error (MSE) =  $\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$



## ChatGPT

Bu görsel, En Küçük Kareler Yöntemi ile parametre tahminini açıklamaktadır.

### En Küçük Kareler Yöntemi (Least Squares Method):

1. "En İyi Uyum" (Best Fit), gerçek  $Y$  değerleri ile tahmin edilen  $Y$  değerleri arasındaki farkların minimum olduğu durumu ifade eder. Pozitif farklar negatif farkları götürdüğü için hataların kareleri alınır.
2. LS (Least Squares), karelenmiş farkların (hataların) toplamını (SSE - Sum of the Squared Errors) minimize eder.
3. Ortalama karelenmiş hata (MSE - Mean Squared Error) aşağıdaki formülle hesaplanır:  
$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Burada  $Y_i$  gerçek değerleri,  $\hat{Y}_i$  ise model tarafından tahmin edilen değerleri ifade eder ve  $n$  gözlem sayısını temsil eder.

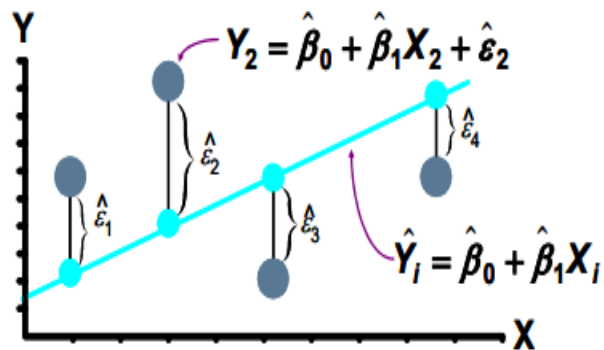
Bu görsel, regresyon analizinde parametre tahmini yapılırken en çok kullanılan yöntemlerden birini, yani en küçük kareler yöntemini anlatan bir ders notu parçasıdır.

Bu yöntem, regresyon çizgisinin veri noktalarına olabildiğince yakın olmasını sağlayarak modelin doğruluğunu artırmayı amaçlar.



## Least Squares Graphically

LS minimizes  $\sum_{i=1}^n \hat{\varepsilon}_i^2 = \hat{\varepsilon}_1^2 + \hat{\varepsilon}_2^2 + \hat{\varepsilon}_3^2 + \hat{\varepsilon}_4^2$



## Interpretation of Coefficients

- 1. Slope ( $\hat{\beta}_1$ )
  - Estimated Y Changes by  $\hat{\beta}_1$  for Each 1 Unit Increase in X
    - If  $\hat{\beta}_1 = 2$ , then Y Is Expected to Increase by 2 for Each 1 Unit Increase in X
- 2. Y-Intercept ( $\hat{\beta}_0$ )
  - Average Value of Y When X = 0
    - If  $\hat{\beta}_0 = 4$ , then Average Y Is Expected to Be 4 When X Is 0



## ChatGPT

Görsel, En Küçük Kareler Yöntemi'nin grafiksel bir temsilini ve regresyon katsayılarının yorumlanmasını içermektedir.

### En Küçük Kareler Grafiği:

- En Küçük Kareler Yöntemi (LS), hata terimlerinin karelerinin toplamını minimize eder:
$$\sum \epsilon_i^2 = \epsilon_1^2 + \epsilon_2^2 + \epsilon_3^2 + \epsilon_4^2$$
- Grafikte, gerçek Y değerleri ve regresyon çizgisinin tahmin ettiği Y değerleri arasındaki dikey mesafeler ( $\hat{\epsilon}_1, \hat{\epsilon}_2$ , vs.), hata terimlerini göstermektedir.
- Denklem olarak,  $\hat{Y} = \beta_0 + \beta_1 X_i + \epsilon_i$  şeklinde ifade edilir.  $\hat{Y}$  tahmin edilen değerleri,  $\beta_0$  Y-kesişimini,  $\beta_1$  ise X'in her bir birim artışı için Y'deki değişimi (eğimi) ifade eder.

### Katsayıların Yorumlanması:

1. Eğim ( $\hat{\beta}_1$ ):
  - Her bir birim X artışı için tahmini Y değişimi  $\hat{\beta}_1$  kadardır.
  - Örneğin,  $\hat{\beta}_1 = 2$  ise, X'in her bir birim artışı için Y'nin 2 birim artması beklenir.
2. Y-kesişimi ( $\hat{\beta}_0$ ):
  - X = 0 iken Y'nin ortalama değeri.
  - Örneğin,  $\hat{\beta}_0 = 4$  ise, X 0 iken ortalama Y'nin 4 olması beklenir.

Bu görsel, doğrusal regresyon modelindeki katsayıların nasıl hesaplanacağını ve bu katsayıların nasıl yorumlanacağını açıklayan bir ders notu parçasıdır.



## Assume that our model is $Y=\beta$

- How can we estimate  $\beta$  using LS?
- Least Squares Minimize squared error

$$\begin{aligned}\sum_{i=1}^n \varepsilon_i^2 &= \sum_{i=1}^n (y_i - \beta)^2 & -2 \sum_{i=1}^n (y_i - \beta) &= 0 \\ \frac{\partial \sum_{i=1}^n (y_i - \beta)^2}{\partial \beta} &= 0 & \sum_{i=1}^n y_i - \sum_{i=1}^n \beta &= 0 \\ & & \sum y &= n\beta \\ & & \beta &= \frac{1}{n} \sum y\end{aligned}$$

## A new look

- A linear first order model (X has d dim.)
- $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_d X_d + \varepsilon$
- This can be written in matrix form as
- $Y_{n \times 1} = X_{n \times (d+1)} \beta_{(d+1) \times 1} + \varepsilon_{n \times 1}$
- n is the sample size



## ChatGPT

Bu görselde, lineer regresyon modelinin en küçük kareler yöntemiyle nasıl tahmin edileceği anlatılmaktadır.

Modelimiz şu şekilde varsayılmış:  $Y = \beta$ . Burada  $Y$  bağımlı değişkeni ve  $\beta$  tahmin edilmek istenen parametreyi ifade ediyor.

En küçük kareler yöntemi (Least Squares - LS), hata terimlerinin karelerinin toplamını minimize ederek parametre tahmininde bulunur. Matematiksel ifadeyle:

$$\sum (e_i)^2 = \sum (Y_i - \beta)^2.$$

Bu ifade, gözlemlenen  $Y$  değerleri ile modelimizin tahmin ettiği  $Y$  değerleri ( $\beta$ ) arasındaki farkların karelerinin toplamını minimize eder. Optimal  $\beta$  değerini bulmak için, bu toplamın  $\beta$ 'ya göre türevi alınır ve sıfıra eşitlenir. Bu işlem sonucunda elde edilen formül şu şekildedir:

$$\beta = (1/n) \sum Y_i.$$

Bu,  $\beta$ 'nin basitçe  $Y$  değerlerinin aritmetik ortalaması olduğunu gösterir.

Görselin ikinci kısmında ise, birden fazla bağımsız değişken içeren bir lineer regresyon modeli verilmiştir. Bu modelde,  $Y$  bir dizi  $X$  bağımsız değişkenine bağlı olarak ifade edilir ve her bir  $X$ 'in kendine ait bir  $\beta$  katsayısı vardır. Bu model şu şekilde gösterilir:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_d X_d + \epsilon.$$

$\epsilon$ , hata terimidir ve modelimizin gözlemlerle tam olarak uyuşmayan kısmını ifade eder. Bu model matris formunda şu şekilde yazılır:

$$Y_{n1} = X_{n(d+1)} \beta_{(d+1)1} + \epsilon_{n1}.$$

Burada  $n$ , örneklem büyüklüğünü ifade eder.

Bu görsel, muhtemelen bir optimizasyon teoremleri veya istatistiksel modelleme dersinin sunum slaytlarından biridir.





## Example-1

- $Y_{n*1} = X_{n*(d+1)} \beta_{(d+1)*1} + \epsilon_{n*1}$
- $n=4, d=1$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{bmatrix} * \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix}$$

$$Y_i = 1 * \beta_0 + \beta_1 * X_i + \epsilon_i$$

## Example-2

- $Y_{n*1} = X_{n*(d+1)} \beta_{(d+1)*1} + \epsilon_{n*1}$
- $n=4, d=2$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ 1 & x_{31} & x_{32} \\ 1 & x_{41} & x_{42} \end{bmatrix} * \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix}$$

$$Y_i = 1 * \beta_0 + \beta_1 * X_{i1} + \beta_2 * X_{i2} + \epsilon_i$$

Bu görselde, iki farklı lineer regresyon model örneği gösterilmektedir. Her iki örnek de modelin matris formülasyonunu kullanarak nasıl ifade edileceğini gösterir.

#### Örnek 1:

Bu örnekte, tek bir bağımsız değişken ( $d=1$ ) için basit lineer regresyon modeli verilmiştir. Örneklem büyüklüğü ( $n$ ) 4'tür. Modelin matris formülasyonu şöyledir:

$$Y = X\beta + \epsilon$$

Burada:

- $Y$  bağımlı değişken vektörüdür ( $Y_1, Y_2, Y_3, Y_4$ ).
- $X$  tasarım matrisidir, her satırda bağımsız değişkenin değerleri ve bir sabit (1) yer alır ( $1, x_1; 1, x_2; 1, x_3; 1, x_4$ ).
- $\beta$  katsayı vektörüdür ( $\beta_0, \beta_1$ ),  $\beta_0$  kesme terimi ve  $\beta_1$  eğim katsayısıdır.
- $\epsilon$  hata terimi vektörüdür ( $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4$ ).

#### Örnek 2:

Bu örnekte, iki bağımsız değişken ( $d=2$ ) için çoklu lineer regresyon modeli verilmiştir. Örneklem büyüklüğü yine 4'tür. Modelin matris formülasyonu şöyledir:

$$Y = X\beta + \epsilon$$

Burada:

- $Y$  bağımlı değişken vektörüdür ( $Y_1, Y_2, Y_3, Y_4$ ).
- $X$  tasarım matrisi, her satırda iki bağımsız değişkenin değerleri ve bir sabit içerir ( $1, x_{11}, x_{12}; 1, x_{21}, x_{22}; 1, x_{31}, x_{32}; 1, x_{41}, x_{42}$ ).
- $\beta$  katsayı vektörüdür ( $\beta_0, \beta_1, \beta_2$ ),  $\beta_0$  kesme terimi,  $\beta_1$  ve  $\beta_2$  ise eğim katsayılarıdır.
- $\epsilon$  hata terimi vektörüdür ( $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4$ ).

Her iki örnekte de, bağımlı değişken  $Y$ 'nin değerleri, bağımsız değişken(ler)in değerleri, katsayılar ve hata terimleri ile ilişkilendirilmiştir. Bu modeller, veri setindeki ilişkileri anlamak ve tahminler yapmak için kullanılır.

### Example-3

$$Y_i = 1 * \beta_0 + \beta_1 * X_i + \beta_2 * X_i^2 + \varepsilon_i$$

- $n=4, d=1, \text{order}=2$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ 1 & x_4 & x_4^2 \end{bmatrix} * \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

- All examples have the following form:
- $Y=X\beta$
- How can we estimate  $\beta$ ?
- $X^{-1}Y=X^{-1}X\beta$  ( $X^{-1}X=I$ )
- $\beta=X^{-1}Y$  (OK, but what if  $X$  is not square matrix?)
- $X^TY=X^TX\beta$  ( $X^TX$  is always a square matrix)
- $(X^TX)^{-1}(X^TY)=(X^TX)^{-1}(X^TX)\beta$  [ $(X^TX)^{-1}(X^TX)=I$ ]
- $\beta=(X^TX)^{-1}(X^TY)$

Görselde üçüncü bir örnek olarak polinom regresyon modeli anlatılmaktadır. Polinom regresyon, bağımsız değişkenin kuvvetlerini kullanarak daha karmaşık ilişkileri modellemeye olanak tanır.

### Örnek 3:

Bu örnekte, bir bağımsız değişken ( $d=1$ ) için ikinci dereceden polinom regresyon modeli verilmiştir. Örneklem büyüklüğü ( $n$ ) 4'tür ve bağımsız değişkenin derecesi (order) 2'dir. Model şu şekildedir:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i$$

Burada:

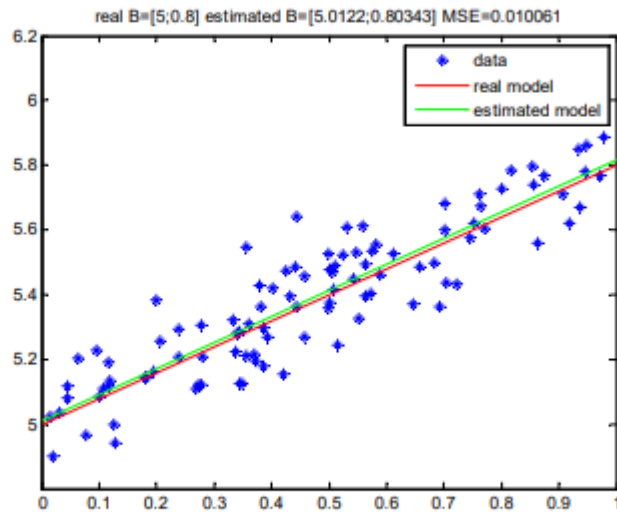
- $Y_i$  bağımlı değişkenin  $i$ 'nci gözlemidir.
- $X_i$  bağımsız değişkenin  $i$ 'nci gözlemidir.
- $X_i^2$  bağımsız değişkenin karesidir, bu non-lineer bir ilişkiyi temsil eder.
- $\beta_0, \beta_1, \beta_2$  modelin katsayılarıdır.
- $\epsilon_i$  hata terimidir.

Görselin alt kısmında,  $\beta$  katsayılarının tahmini için kullanılan matris cebiri anlatılmaktadır. Genel model  $Y = X\beta$  şeklinde ifade edilir ve  $\beta$  katsayıları şu şekilde tahmin edilir:

- $X'Y = X'X\beta$
- $(X'X)^{-1}X'Y = (X'X)^{-1}X'X\beta$
- $\beta = (X'X)^{-1}X'Y$

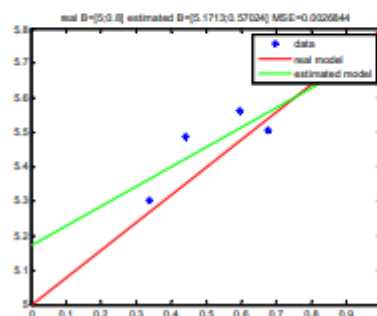
Bu adımlar, en küçük kareler tahmincisinin formülasyonudur ve  $\beta$  katsayılarını bulmak için kullanılır.  $X'$  matrisin transpozisini,  $(X'X)^{-1}$  ise  $X'X$  matrisinin tersini ifade eder. Bu formülasyon,  $X$  matrisi kare olmasa bile uygulanabilir çünkü  $X'X$  her zaman kare bir matristir ve tersi alınabilir (eğer  $X$  tam sıralıysa). Bu yöntem, verilen bağımsız değişkenlerle bağımlı değişken arasındaki ilişkiyi modellemek için katsayıları tahmin etmek amacıyla kullanılır.

## Lin\_reg1.m

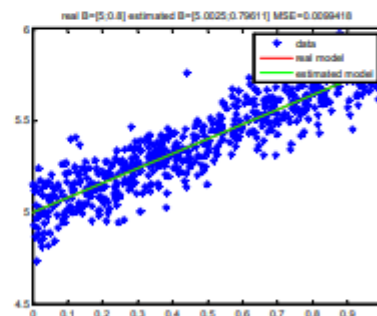


## Sample size effect lin\_reg1.m

N=6



N=600





### ChatGPT

Bu görselde lineer regresyon analizinin sonuçlarını gösteren grafikler yer almaktadır.

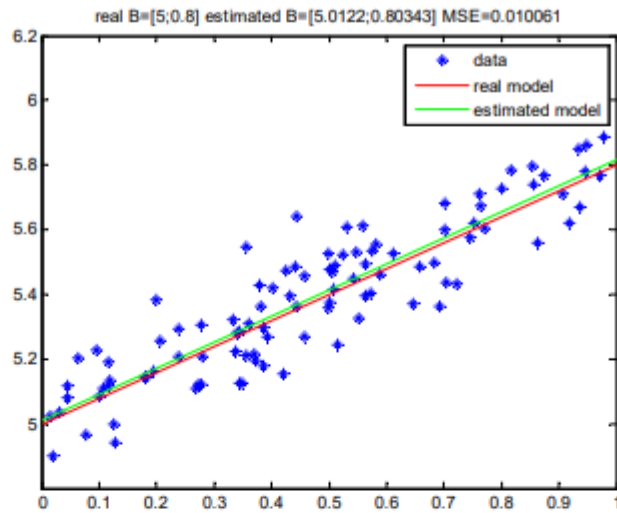
Üstteki grafik, "Lin\_reg1.m" başlıklı, bir lineer regresyon modelinin veri seti üzerindeki uygulamasını göstermektedir. Grafikte mavi noktalar gerçek veri noktalarını, yeşil çizgi gerçek modeli ( $\beta = [5 \ 0.5]$  olan), ve kırmızı çizgi tahmin edilen modeli temsil etmektedir. Tahmin edilen modelin katsayıları  $\beta = [5.0122, 0.80343]$  olarak belirtilmiştir ve modelin ortalama karesel hatası (MSE) 0.010061 olarak hesaplanmıştır. Bu, modelin veriye ne kadar iyi uyduğunu gösteren bir ölçüttür; değer ne kadar düşükse model o kadar iyi uyarlanmış demektir.

Alt grafik, örneklem büyüklüğünün (sample size) tahmin edilen model üzerindeki etkisini göstermektedir. Sol taraftaki küçük alt grafikte,  $N=6$  (örneklem büyüklüğü 6) için bir lineer regresyon modeli gösterilmekte ve sağdaki büyük alt grafikte,  $N=600$  için aynı modelin uygulaması yer almaktadır. Sol taraftaki grafikte, az sayıda veri noktası olduğu için modelin veriye uyumu daha az kesin görünmektedir ve bu da daha yüksek bir MSE (0.026424) değerine yol açmaktadır. Sağ taraftaki grafikte, çok daha fazla veri noktası ile modelin tahminleri gerçek verilere daha yakın olmakta ve bu da daha düşük bir MSE (0.010263) değeri ile sonuçlanmaktadır. Bu grafikler, daha fazla veri noktasının, modelin verileri daha iyi temsil etmesine ve dolayısıyla daha iyi bir model uyumu sağlamasına yardımcı olduğunu göstermektedir.



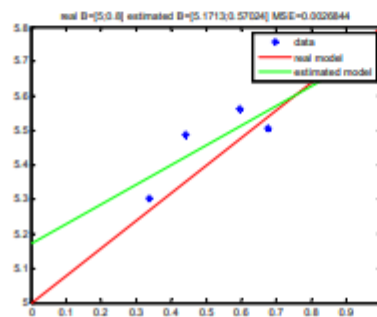


## Lin\_reg1.m

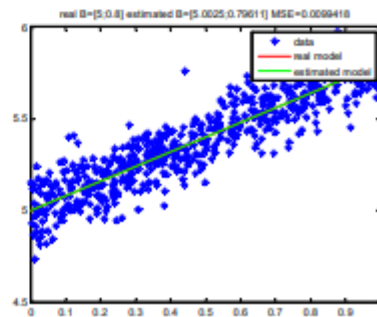


## Sample size effect lin\_reg1.m

N=6



N=600





## ChatGPT

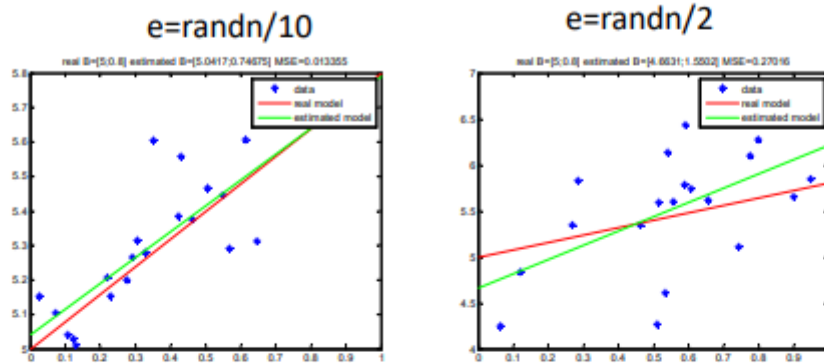
Bu görsel, lineer regresyon modelinin iki farklı örneklem büyüklüğünde uygulamasını göstermektedir.

Üst grafikte "Lin\_reg1.m" adı verilen bir lineer regresyon modelinin sonuçları yer almakta. Burada, gerçek modelin parametreleri  $\beta_0=5$  ve  $\beta_1=0.5$  iken, tahmin edilen parametreler  $\beta_0=5.0122$  ve  $\beta_1=0.80343$  olarak bulunmuş. Modelin ortalama karesel hatası (MSE) ise 0.010061 olarak hesaplanmış. Mavi noktalar gerçek veri noktalarını, yeşil çizgi gerçek modeli ve kırmızı çizgi ise tahmin edilen modeli temsil etmektedir.

Alt grafiklerde ise örneklem büyüklüğünün (sample size) model tahmini üzerindeki etkisini gösterilmektedir. Sol tarafta,  $N=6$  için yapılan tahminde, gerçek model yeşil, tahmin edilen model kırmızı ve veri noktaları mavi olarak gösterilmiş. Sağ tarafta ise  $N=600$  için yapılan tahminde, aynı renk kodlaması kullanılarak, çok daha fazla veri noktası ile daha düşük bir MSE (0.010263) elde edilmiştir. Örneklem büyüklüğünün artmasıyla birlikte modelin verileri daha iyi temsil ettiği ve tahminlerin daha doğru olduğu gözlemlenmektedir.

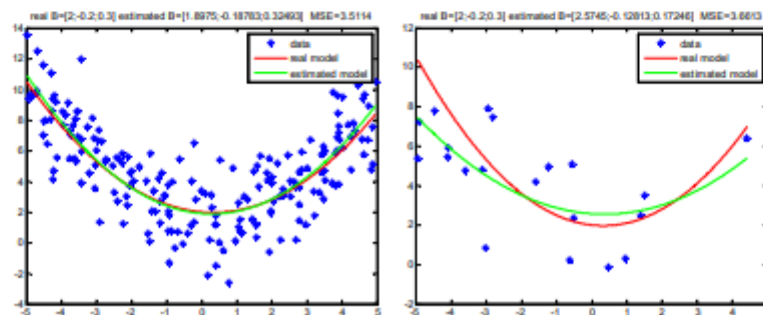


## Error rate effect lin\_reg1.m



$$Y = b_0 + b_1 * x + b_2 * x^2$$

## Lin\_reg2.m



Bu görselde, hata oranının ve polinom derecesinin lineer regresyon model tahminleri üzerindeki etkisini gösteren dört grafik bulunmaktadır.

#### Üst Grafikler - Hata Oranı Etkisi (Error rate effect lin\_reg1.m):

Bu iki grafik, hata terimlerinin büyüklüğünün tahmin edilen modele etkisini karşılaştırmaktadır. Her iki grafikte de, tahmin edilen modelin katsayıları ve modelin ortalama karesel hatası (MSE) gösterilmiştir.

- Sol grafikte, hata terimi  $e = \frac{\text{randn}}{10}$  ile ifade edilmiştir. Bu, hataların normal dağılımı takip ettiğini ve ondalık bir faktörle ölçeklendirildiğini gösterir. Bu durumda, tahmin edilen modelin katsayıları  $\beta_0$  ve  $\beta_1$ , gerçek değerlere oldukça yakın bulunmuş ve MSE 0.003355 olarak hesaplanmış.
- Sağ grafikte, hata terimi  $e = \frac{\text{randn}}{2}$  olarak belirlenmiş. Hataların büyüklüğünün artırılması modelin tahmin kabiliyetini olumsuz etkilemiş ve MSE 0.267016 olarak daha yüksek bulunmuştur.

#### Alt Grafikler - Polinom Derecesi Etkisi ( $Y=b_0+b_1x+b_2x^2$ Lin\_reg2.m):

Bu iki grafik, polinom derecesinin model tahminleri üzerindeki etkisini göstermektedir. Grafikler, bağımsız değişkenin (x) ve onun karesinin ( $x^2$ ) kullanıldığı ikinci dereceden bir polinom regresyon modelini temsil eder.

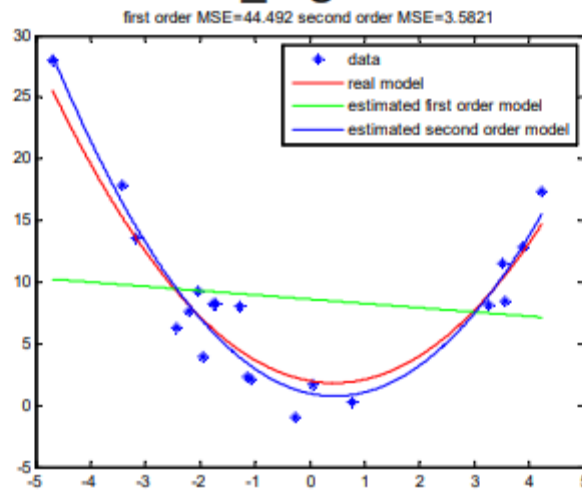
- Sol grafik, verilerin daha düşük bir polinom derecesi ile modellemesi sonucunu gösterir. Tahmin edilen modelin katsayıları ve MSE'si gösterilmiştir. Model, verileri oldukça iyi tahmin etmiş gibi görünmektedir, ancak bazı sapmalar mevcuttur (MSE=0.514).

- Sağ grafik, verilerin daha yüksek bir polinom derecesi ile modellemesi sonucunu gösterir. Bu modelde de tahmin edilen katsayılar ve MSE değeri verilmiştir. Model, özellikle veri noktalarının yoğun olduğu bölgelerde daha iyi bir uyum sağlamaktadır (MSE=0.1603).

Her iki durumda da, modelin veriye uyumu (yeşil çizgi), tahmin edilen modelin uyumu (kırmızı çizgi) ve gerçek veri noktaları (mavi noktalar) grafikte görselleştirilmiştir. Hata oranının ve modelin karmaşıklığının (polinom derecesi) modelin performansını nasıl etkilediği bu grafiklerle açıkça gösterilmektedir. Hata oranının artması ve modelin karmaşıklığının azalması MSE'yi artırırken, uygun bir model karmaşıklığı seçimi modelin veriyi daha iyi temsil etmesini ve dolayısıyla daha düşük bir MSE değeri elde edilmesini sağlar.

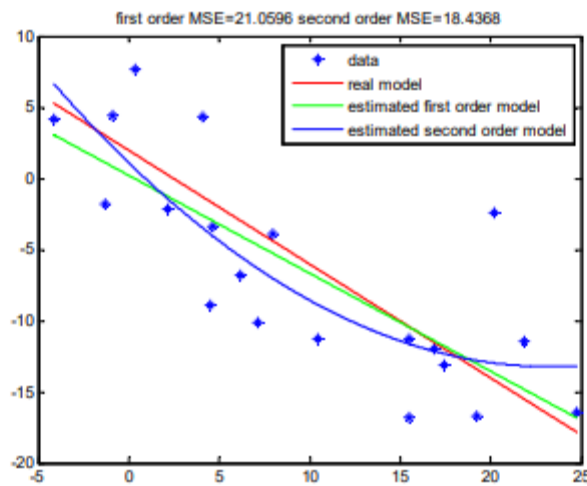
## Real 2<sup>nd</sup> order, estimated 1<sup>st</sup> and 2<sup>nd</sup> order

### Lin\_reg3.m



## Real 1<sup>st</sup> order, estimated 1<sup>st</sup> and 2<sup>nd</sup> order

### Lin\_reg4.m



← What happens in these areas? →



## ChatGPT

Bu görselde, iki farklı lineer regresyon modelinin karşılaştırmalarını içeren iki grafik bulunmaktadır.

### **Üst Grafik - Gerçek 2. Derece, Tahmini 1. ve 2. Derece (Lin\_reg3.m):**

Grafik, gerçek bir ikinci dereceden polinom veri seti üzerinde tahmini birinci ve ikinci dereceden polinom modellerin uyumunu göstermektedir. Mavi noktalar gerçek veri noktalarını, yeşil çizgi tahmini birinci dereceden modeli (lineer model), ve kırmızı çizgi tahmini ikinci dereceden modeli (polinom model) temsil etmektedir. İkinci dereceden model, verilerin dağılımını birinci dereceden modele göre daha iyi yakalamış ve daha düşük bir ortalama karesel hata (MSE) değeri olan 3.5821 ile sonuçlanmıştır. Birinci derecenin MSE'si ise 44.492 olarak çok daha yüksektir, bu da modelin verilere iyi uymadığını gösterir.

### **Alt Grafik - Gerçek 1. Derece, Tahmini 1. ve 2. Derece (Lin\_reg4.m):**

Bu grafik, gerçek birinci dereceden (lineer) bir veri seti üzerinde hem tahmini birinci dereceden hem de ikinci dereceden modellerin uyumunu gösterir. Yine mavi noktalar gerçek veri noktalarını, yeşil çizgi tahmini birinci dereceden modeli, ve kırmızı çizgi tahmini ikinci dereceden modeli ifade eder. Bu durumda, birinci dereceden modelin verilere uyumu daha iyi olup MSE'si 2.01568'dir. İkinci dereceden modelin ise MSE'si 18.4368 olarak hesaplanmıştır ki bu, modelin veri setine uyumunun daha kötü olduğunu gösterir.

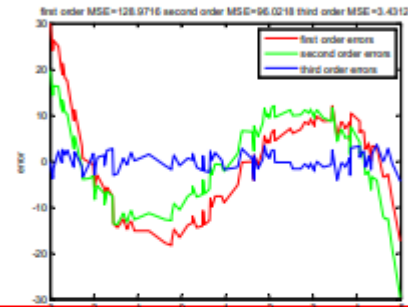
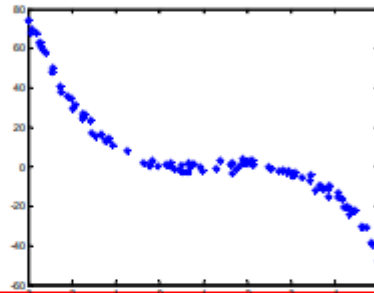
Grafiklerin altında yer alan kırmızı oklarla belirtilen "What happens in these areas?" notu, her iki modelin de veri setinin uç noktalarında (grafiklerin kenarlarında) nasıl performans gösterdiğine dikkat çekmektedir. Bu alanlar, modelin ekstrapolasyon kabiliyetinin bir göstergesi olabilir; yani modelin, eğitim veri setinde yer almayan yeni veri noktaları için ne kadar iyi tahminler yapabileceğini anlamak için incelenebilir.



```
% real model Y=B0+B1*X+B2*X^2+B3*X^3
% estimated model1 Y=B0+B1*X
% estimated model2 Y=B0+B1*X+B2*X^2
% estimated model3 Y=B0+B1*X+B2*X^2+B3*X^3
```

```
B = [ 1 -4 4 -1]'
fBhat = [ 16.3823 -9.4018 ]'
sBhat = [ 10.3682 -11.4712 1.1441 ]'
tBhat = [ 1.0170 -4.0199 3.9625 -0.9901 ]'
```

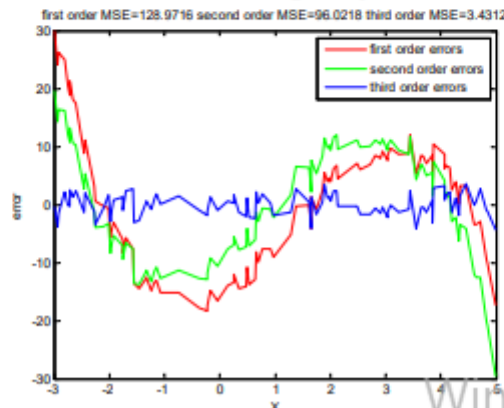
Lin\_reg5.m



Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG. DEPT.

## X-error graph



If there is a pattern in X-error graph, there is something unmodeled.

If the model degree is sufficient,  
the errors have the same variances along X.

Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG. DEPT.

## ChatGPT

Bu görselde, üç farklı polinom derecesi için yapılan lineer regresyon model tahminlerinin sonuçları ve hatalarının grafiksel analizi yer almaktadır.

### Model Bilgileri:

- Gerçek model:  $Y = B_0 + B_1 * X + B_2 * X^2 + B_3 * X^3$
- Tahmini model 1:  $Y = B_0 + B_1 * X$
- Tahmini model 2:  $Y = B_0 + B_1 * X + B_2 * X^2$
- Tahmini model 3:  $Y = B_0 + B_1 * X + B_2 * X^2 + B_3 * X^3$

Gerçek modelin katsayıları  $B = [1, -4, 4, -1]$  olarak verilmiş, farklı model tahminlerinin katsayıları ise sırasıyla  $fBhat$ ,  $sBhat$ , ve  $tBhat$  olarak gösterilmiştir. Bu katsayılar tahmin edilen modellerin ne kadar iyi performans gösterdiğini göstermektedir.

### Üstteki Sol Grafik:

Gerçek veri noktalarının dağılımını göstermektedir. Bu grafik, verilerin gerçek dünya dağılımını temsil eder.

### Üstteki Sağ Grafik:

Üç farklı modelin hata terimlerini (tahmin edilen değerler ile gerçek değerler arasındaki farklar) karşılaştırmaktadır. Her renk, farklı bir modelin hatalarını temsil eder. Grafik, her bir modelin hangi noktalarda daha iyi veya daha kötü performans gösterdiğini göstermek için hata büyüklüklerini X eksenini boyunca çizer.

- Kırmızı çizgi birinci dereceden modelin hatalarını,
- Yeşil çizgi ikinci dereceden modelin hatalarını,
- Mavi çizgi üçüncü dereceden modelin hatalarını temsil eder.

### Alt Grafik (X-error graph):

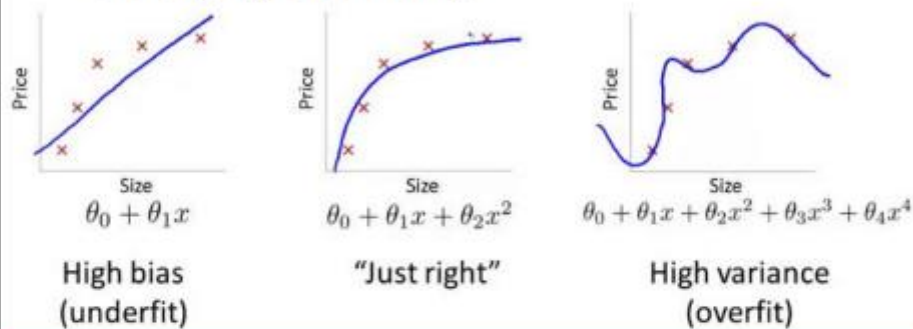
Bu grafik, X eksenini boyunca hataların dağılımını gösterir. Eğer grafikte bir desen (pattern) varsa, bu, modelde modellenmemiş bir şey olduğunu gösterir. Eğer model derecesi yeterliyse ve hatalar X boyunca sabit varyanslara sahipse, bu modelin verileri iyi temsil ettiğini gösterir. Grafikte farklı model derecelerinin hatalarının nasıl değiştiği gösterilmiştir.

Hataların grafikte herhangi bir sistematik desen oluşturmaması, modelin veri setindeki örüntüleri iyi yakaladığını gösterir. Eğer hataların dağılımında belirgin bir desen varsa veya hataların varyansı X boyunca değişiyorsa, bu durum modelin eksik olduğuna işaret edebilir (örneğin, yeterince yüksek dereceden bir polinom kullanılmamış olabilir).

Özetle, görseldeki grafikler ve tablolar, farklı polinom derecelerinin veriye nasıl uyduğunu ve hataların nasıl dağıldığını analiz etmek için kullanılır. Bu analiz, model seçiminde ve modelin doğruluğunu değerlendirirken önemlidir.

# Overfitting

- **Overfitting** occurs when a statistical model describes random error or noise instead of the underlying relationship.

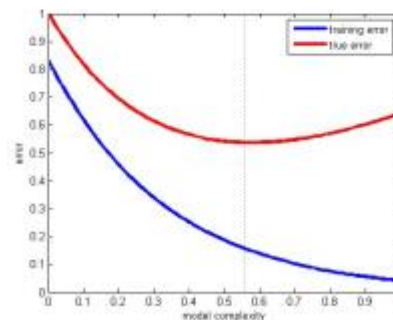


Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG. DEPT.

# Model Validation

- Training set MSE is not reliable. **WHY?**
- **Because, we can not determine the overfitting with training set MSE.**
- Training set is used for parameter estimation.
- Test set is used for model validation.



Windows'u Etk  
Windows'u etkinlesti

### **Overfitting (Aşırı Uyum):**

Aşırı uyum, istatistiksel bir modelin altında yatan ilişkiden ziyade rastgele hata veya gürültüyü tanımladığı durumu ifade eder. Görselde üç farklı grafik gösterilmekte:

- Sol grafik, yüksek yanlılık (high bias) gösteren ve genellikle "underfitting" (yetersiz uyum) olarak adlandırılan bir durumu temsil ediyor. Model, verilerin altında yatan trendi yakalamada yetersiz kalıyor.
- Ortadaki grafik, modelin verilere "tam olarak uyduğu" ideal durumu gösteriyor. Model, verilerin genel eğilimini yakalıyor ve aynı zamanda rastgele varyasyonları aşırı öğrenmiyor.
- Sağdaki grafik ise aşırı uyumu (overfitting) temsil ediyor. Model, verilerin içindeki rastgele gürültüyü veya hataları öğrenmiş ve bunu genel trend olarak kabul etmiş, bu da modelin yeni verilere genelleme yapma kabiliyetini düşürür.

### **Model Doğrulama (Model Validation):**

Model doğrulama kısmında ise, eğitim veri setinin ortalama karesel hatasının (MSE) güvenilir olmadığı ve bunun nedeninin aşırı uyumun eğitim seti MSE ile belirlenememesi olduğu belirtilmiş. Eğitim seti, model parametrelerini tahmin etmek için kullanılırken, test seti modelin doğruluğunu değerlendirmek için kullanılır. Bu aşamada modelin gerçek dünya verileri üzerinde ne kadar iyi performans gösterdiği test edilir.

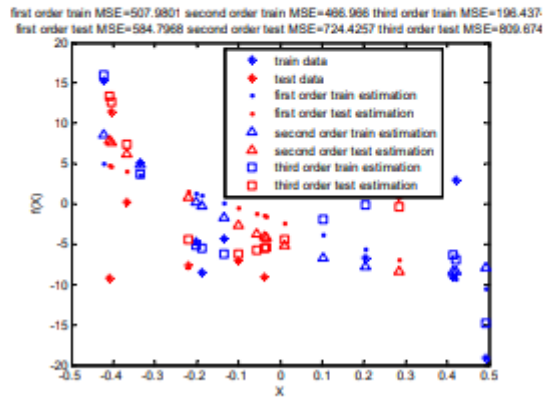
Görselin alt kısmında yer alan küçük grafikte, eğitim ve test hatalarının farklı model karmaşıklıklarına göre nasıl değiştiğini gösteren bir eğri bulunmaktadır. Eğitim hatası genellikle modelin karmaşıklığı arttıkça azalırken, test hatası belirli bir noktadan sonra artmaya başlar. Bu nokta genellikle modelin aşırı uyuma başladığı noktadır.

Özetle, bu görselde verilen bilgiler, modelleme yaparken doğru model karmaşıklığını seçmenin ve modelin gerçek dünya verileri üzerindeki performansını doğru bir şekilde değerlendirmenin önemini vurgulamaktadır.



## Model Validation

- Training and test sets are separated data samples.
- Lin\_reg6.m



## $Y=X*\beta$ Linear System Construction

- Data=  $(X_{n \times d}, Y_{n \times 1})$
- n: number of data, d: dimension of data
- Model  $Y= \beta_1 + \beta_2 * X$
- $X_{n \times 2} = [1_{n \times 1} \ X_{n \times 1}]$   $\beta_{2 \times 1} = [\beta_1 ; \beta_2]$
- Model  $Y= \beta_1 * X + \beta_2 * X^2$
- $X_{n \times 2} = [X_{n \times 1} \ X_{n \times 1}^2]$   $\beta_{2 \times 1} = [\beta_1 ; \beta_2]$
- Model  $Y= \beta_1 + \beta_2 * X^2 + \beta_3 * X^3$
- $X_{n \times 3} = [1_{n \times 1} \ X_{n \times 1}^2 \ X_{n \times 1}^3]$   $\beta_{3 \times 1} = [\beta_1 ; \beta_2 ; \beta_3]$





## ChatGPT

Bu görselde, model doğrulama süreçleri ve lineer sistem kurulumu üzerine bilgiler yer almaktadır.

### Model Doğrulama:

- Eğitim ve test setleri ayrı veri örnekleridir. Bir modeli değerlendirirken, modelin öğrenmesi için kullanılan eğitim setinin yanı sıra, modelin genellebilirliğini test etmek için ayrı bir test seti kullanılır.
- Görseldeki grafikte "Lin\_reg6.m" adlı bir dosyada yapılan model tahminlerinin sonuçları gösteriliyor. Farklı şekil ve renklerdeki işaretler, eğitim ve test veri setlerini ve bu setler üzerinde yapılan tahminlerin derecelerini (birinci, ikinci ve üçüncü derece polinomlar) temsil ediyor.

### Lineer Sistem Kurulumu:

- Veri,  $(X^n, Y^n)$  olarak ifade edilmiş. Burada  $n$ , veri noktalarının sayısını ve  $d$ , veri noktalarının boyutunu (bağımsız değişkenlerin sayısı) belirtir.
- İlk model  $Y = \beta_1 + \beta_2 * X$  olarak ifade edilmiş ve bu model için gerekli olan  $X_{n \times 2}$  matris ve  $\beta_{2 \times 1}$  vektörü belirtilmiştir.
- İkinci model  $Y = \beta_1 * X + \beta_2 * X^2$  ve bu model için de  $X_{n \times 2}$  ve  $\beta_{2 \times 1}$  vektörleri verilmiş.
- Üçüncü model ise  $Y = \beta_1 + \beta_2 * X + \beta_3 * X^3$  ve buna uygun  $X_{n \times 3}$  ve  $\beta_{3 \times 1}$  vektörleri gösterilmiş.

Bu bilgiler ışığında, görseldeki içerik modelin karmaşıklığının arttıkça (polinom derecesinin yükselmesiyle) eğitim seti üzerindeki hatanın genellikle azaldığını, ancak bu durumun her zaman test seti üzerinde iyi genellebilirlik anlamına gelmediğini göstermektedir. Bu nedenle, model karmaşıklığı ve veri setlerinin doğru bir şekilde seçilmesi ve kullanılması önemlidir.

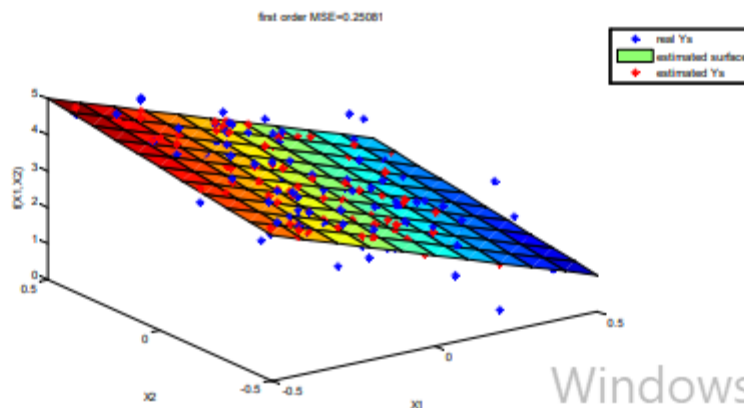
— — — —



## Y=X\*β Linear System Construction

- Model  $Y = \beta_1 * X + \beta_2 * \cos(X^2)$
- $X_{n+2} = [X_{n+1} \cos(X^2)_{n+1}]$   $\beta_{2+1} = [\beta_1 ; \beta_2]$
- Model  $Y = \beta_1 * X_1 + \beta_2 * \cos(X_2^2) + \beta_3 * \sin(X_1)$
- $X_{n+3} = [X_{1n+1} \cos(X_2^2)_{n+1} \sin(X_1)_{n+1}]$   $\beta_{3+1} = [\beta_1 ; \beta_2 ; \beta_3]$
- Model  $Y = \beta_1 + \beta_2 * X_1 * X_2 * X_3 + \beta_3 * X_1$
- $X_{n+3} = [1_{n+1} X_1 * X_2 * X_3_{n+1} X_1_{n+1}]$   $\beta_{3+1} = [\beta_1 ; \beta_2 ; \beta_3]$

## model $Y = B_0 + B_1 * X_1 + B_2 * X_2$ multivariate first order



$$B = [3 \quad -3 \quad 1]'$$

$$\hat{B} = [3.0216 \quad -3.0439 \quad 1.0387]'$$



## ChatGPT

Gönderdiğiniz görüntü, doğrusal sistemlerin yapılandırılmasıyla ilgili bir matematiksel modelleme konusuna ait görünüyor. Metinde, farklı matematiksel ifadeler kullanılarak oluşturulan doğrusal olmayan modeller ve bunların lineer hale getirilmiş halleri sunuluyor. Aynı zamanda, bu modellerin parametre tahminleri ve bunların nasıl hesaplandığı ile ilgili bir grafik de mevcut.

İlk model, " $Y = \beta_1 X_1 + \beta_2 \cos(X_2)$ " şeklinde bir doğrusal olmayan model. Bu modelde, Y'nin  $X_1$  ve  $X_2$ 'nin kosinüsüne bağlı olduğu belirtiliyor. Modelin lineer hale getirilmiş hali ise " $X_{n2} = [X_{n1} \cos(X_2)]$ " matrisi ve " $\beta = [\beta_1; \beta_2]$ " vektörü kullanılarak ifade edilmiş.

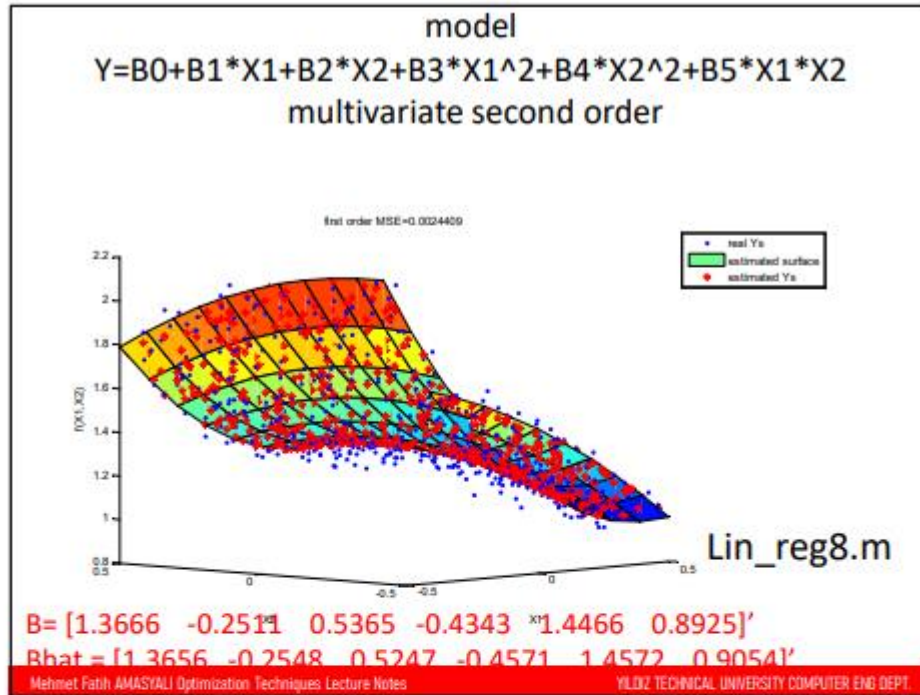
İkinci model, " $Y = \beta_1 X_1 + \beta_2 \cos(X_2) + \beta_3 \sin(X_1)$ " şeklinde üç terimli bir model. Burada, Y'nin  $X_1$ ,  $X_2$ 'nin kosinüsü ve  $X_1$ 'in sinüsüne bağlı olduğu gösteriliyor. Bu modelin lineerleştirilmiş hali ise " $X_{n3} = [X_{n1} \cos(X_2) \sin(X_1)]$ " matrisi ve " $\beta = [\beta_1; \beta_2; \beta_3]$ " vektörü ile temsil ediliyor.

Üçüncü modelde ise, " $Y = \beta_1 + \beta_2 X_1 X_2 X_3 + \beta_3 X_1$ " bir başka doğrusal olmayan model sunulmuş ve bunun lineerleştirilmiş hali " $X_{n3} = [1 \ X_1 X_2 X_3 \ X_1]$ " matrisi ve " $\beta = [\beta_1; \beta_2; \beta_3]$ " vektörü ile ifade edilmiş.

Son kısımda ise, "model  $Y = B_0 + B_1 X_1 + B_2 X_2$ " şeklinde çok değişkenli bir birinci dereceden model ve bu modele ait bir yüzey grafiği bulunuyor. Grafiğin altında, model parametrelerinin tahmini değerleri verilmiş: " $B = [3 \ -3 \ 1]$ " ve tahmini değerlerin daha hassas bir şekilde hesaplanmış hali olan " $B_{hat} = [3.0216 \ -3.0439 \ 1.0387]$ ".

Bu görsel, bir mühendislik veya istatistik dersinde kullanılabilecek, doğrusal olmayan sistemlerin nasıl lineer hale getirilebileceğini ve parametre tahminlerinin nasıl yapılacağını gösteren bir öğretim materyali gibi duruyor.





## What if we can not write linear equation system?

- (linear according to b)
- $Y=b_1*\sin(X)$
- $Y=X_1*b_1*\sin(X_2)$
- $Y=b_1*\sin(X) + b_2*\cos(X)$
- $Y=b_1*\exp(X) + b_2*\cos(X)$
- $Y=b_1*\exp(X_1) + b_2*\cos(X_2) + b_3*\cos(X_1)$
- (non-linear according to b)
- $Y=\sin(b_1*X)$
- $Y=b_1+\sin(b_2*X)$
- $Y=\exp(b_1*X)$
- $Y=(b_1*X)/(b_2+X)$
- $Y=b_1*(\exp(-(X-b_2)^2/(b_3^2)))$

Windows'u Etkir



## ChatGPT

Gönderdiğiniz görselde, çok değişkenli ikinci dereceden bir doğrusal olmayan model ve bu modelin lineer hale getirilmesiyle ilgili bilgiler yer alıyor. Görselde ayrıca, modelin tahmini parametre değerleri ve bu değerlerin daha hassas hesaplanmış halleri (Bhat) ile birlikte bir yüzey grafiği bulunuyor.

Model şu şekilde ifade ediliyor:

$$Y = B_0 + B_1 * X_1 + B_2 * X_2 + B_3 * X_1^2 + B_4 * X_2^2 + B_5 * X_1 * X_2$$

Bu model, bir yanıt değişkeni  $Y$ 'nin, bağımsız değişkenler  $X_1$  ve  $X_2$ , bunların kareleri ve birbiriyle çarpımları arasındaki ilişkiyi ikinci dereceden bir polinom olarak tanımlıyor.

Modelin parametre tahminleri ve daha hassas tahminleri şöyle verilmiş:

$$B = [1.3666, -0.251, 0.5365, -0.4343, -1.4466, 0.8925]'$$

$$\hat{B} = [1.3656, -0.2548, 0.5327, -0.4571, 1.4577, 0.9054]'$$

Ayrıca, görselde lineer olmayan sistemlerde kullanılabilecek alternatif yaklaşımlar sunulmuş. Eğer doğrusal bir denklem sistemi yazılamıyorsa, lineer olmayan fonksiyonlar (sinüs, eşponansiyel vs.) kullanarak ifade edilen modeller öneriliyor. Örneğin:

- Lineer yaklaşıma göre bir model:  $Y = b_1 * \sin(X)$
- Lineer olmayan yaklaşıma göre bir model:  $Y = \sin(b_1 * X)$

Son olarak, görselin alt kısmında bir uyarı notu bulunuyor: "Lineer denklem sistemi yazamazsak ne olur?" ve bu durum için çeşitli lineer ve lineer olmayan denklem örnekleri sunuluyor.

Bu tür modeller genellikle mühendislik, ekonometri ve istatistik alanlarında, veriler arasındaki ilişkileri modellemek ve tahminlerde bulunmak için kullanılır.

## Non-linearity

- A parameter  $\beta$  of the function  $f$  appears nonlinearly if the derivative  $\partial f / \partial \beta$  is a function of  $\beta$ .
- The model  $M(\beta, x)$  is nonlinear if at least one of the parameters in  $\beta$  appear nonlinearly.
- $f(x) = \beta \sin(x)$ ,  $\partial f / \partial \beta = \sin(x)$ , which is independent of  $\beta$ , so the model  $M(\beta, x)$  is linear.
- $f(x) = \sin(\beta x)$ ,  $\partial f / \partial \beta = x \cos(\beta x)$ , which is dependent of  $\beta$ , so the model  $M(\beta, x)$  is non-linear.

## Non-linearity

- $f(x) = \beta_1 \sin(x) + \beta_2 \cos(x)$ ,  $\partial f / \partial \beta_1 = \sin(x)$ , which is independent of  $\beta_1$ ,  $\partial f / \partial \beta_2 = \cos(x)$ , which is independent of  $\beta_2$ , so the model  $M(\beta, x)$  is linear.
- $f(x) = \beta_1 + \cos(\beta_2 x)$ ,  $\partial f / \partial \beta_1 = 1$ , which is independent of  $\beta_1$ , but  $\partial f / \partial \beta_2 = -x \sin(\beta_2 x)$ , which is dependent of  $\beta_2$ , so the model  $M(\beta, x)$  is non-linear.



data data can you make



### ChatGPT

Görselde, bir fonksiyonun lineer olup olmadığını belirleme üzerine bir açıklama bulunuyor. Bir fonksiyonun veya modelin lineer olup olmadığı, parametrelerin ( $\beta$ ) fonksiyon içindeki davranışına göre değişir.

- Eğer bir fonksiyonun  $\beta$  parametresinin türevi  $\beta$ 'ya bağlı değilse, o zaman bu model lineerdir. Örneğin:  $f(x) = \beta * \sin(x)$ , bu durumda türev  $\frac{\partial f}{\partial \beta} = \sin(x)$  olup  $\beta$ 'ya bağlı değildir ve model lineerdir.
- Eğer  $\beta$  parametresinin türevi  $\beta$ 'ya bağlı ise, model lineer değildir. Örneğin:  $f(x) = \sin(\beta * x)$ , bu durumda türev  $\frac{\partial f}{\partial \beta} = x * \cos(\beta * x)$  olup  $\beta$ 'ya bağlıdır ve model lineer değildir.

Görseldeki diğer örneklerde de benzer şekilde, türevin  $\beta$  parametresine bağımlılığına göre modellerin lineerliği değerlendirilmiş. İlk örnekte

$f(x) = \beta_1 * \sin(x) + \beta_2 * \cos(x)$  fonksiyonunun her iki türevi de ( $\frac{\partial f}{\partial \beta_1}$  ve  $\frac{\partial f}{\partial \beta_2}$ )  $\beta$ 'lara bağlı değildir ve bu nedenle model lineer kabul edilir.

İkinci örnekte  $f(x) = \beta_1 + \cos(\beta_2 * x)$  fonksiyonunun ilk türevi ( $\frac{\partial f}{\partial \beta_1}$ )  $\beta_1$ 'e bağlı değilken, ikinci türevi ( $\frac{\partial f}{\partial \beta_2}$ )  $\beta_2$ 'ye bağlıdır. Bu nedenle model lineer değildir.

Bu bilgiler, genellikle matematiksel modelleme ve istatistiksel analizde önemli olan fonksiyonların doğası hakkında temel bir anlayış sağlar. Lineer modeller analitik çözüm bulma ve yorumlama açısından genellikle daha basitken, lineer olmayan modeller daha karmaşık ilişkileri modelleyebilir ancak analizleri ve çözümleri daha zordur.



## What if we can not write linear equation system?

- There are two ways:
  - Transformations to achieve linearity
  - Nonlinear regression (iterative estimation)

## Transformations to achieve linearity

- Some tips:
- $\ln(e)=1, \ln(1)=0$
- $\ln(x^r)=r*\ln(x)$
- $\ln(e^A)=A*\ln(e)=A$
- $\ln(A*B)=\ln(A)+\ln(B)$
- $\ln(A/B)=\ln(A)-\ln(B)$
- $e^{(A*B)}=(e^A)^B$
- $e^{(A+B)}=(e^A)*(e^B)$
- $e^{(A-B)}=(e^A)/(e^B)$

bana bunu türkçe anlat



## ChatGPT

Görselde, doğrusal olmayan bir denklem sistemiyle karşılaşıldığında izlenebilecek iki yol öneriliyor:

1. Lineerliği sağlamak için dönüşümler yapılabilir.
2. Doğrusal olmayan regresyon (iteratif tahmin) kullanılabilir.

Daha sonra, lineerliğe ulaşmak için bazı logaritmik ve üstel işlemlerle ilgili ipuçları veriliyor:

- $\ln(e) = 1, \ln(1) = 0$
- Bir sayının kuvveti için,  $\ln(x^r) = r * \ln(x)$
- Üstel fonksiyonların logaritması,  $\ln(e^A) = A * \ln(e) = A$
- Çarpımın logaritması,  $\ln(A * B) = \ln(A) + \ln(B)$
- Oranın logaritması,  $\ln(A/B) = \ln(A) - \ln(B)$
- Üstel fonksiyonların çarpımı,  $e^{(A+B)} = e^A * e^B$
- Üstel fonksiyonların oranı,  $e^{(A-B)} = \frac{e^A}{e^B}$

Bu dönüşümler, doğrusal olmayan ilişkiler içeren veriler üzerinde lineer regresyon modeli uygulamak için kullanılır. Örneğin, bir veri seti üstel bir trend gösteriyorsa, verilerin logaritmasını almak, bir lineer modelin uygulanabilmesi için verileri dönüştürmektedir. Bu yöntemler, istatistik ve veri biliminde sıkça kullanılır.



## Transformations to achieve linearity

- Original  $Y=b_0 \cdot \exp(b_1 \cdot X)$
- $\ln(Y) = \ln(b_0) + (b_1 \cdot X)$
- $Z = \ln(Y)$ ,  $b_2 = \ln(b_0)$
- $Z = b_2 + b_1 \cdot X$  (linear)
  
- Original  $Y = \exp(b_0) \cdot \exp(b_1 \cdot X)$
- $\ln(Y) = b_0 + b_1 \cdot X$
- $Z = \ln(Y)$
- $Z = b_0 + b_1 \cdot X$  (linear)

## Transformations to achieve linearity

- Original  $Y = (b_0 + b_1 \cdot X)^2$
- $\sqrt{Y} = b_0 + b_1 \cdot X$
- $Z = \sqrt{Y}$ ,
- $Z = b_0 + b_1 \cdot X$  (linear)
  
- Original  $Y = 1 / (b_0 + b_1 \cdot X)$
- $1/Y = b_0 + b_1 \cdot X$
- $Z = 1/Y$
- $Z = b_0 + b_1 \cdot X$  (linear)



### ChatGPT

Gönderdiğiniz görselde, doğrusal olmayan bazı modeller ve bu modelleri doğrusal hale getirebilmek için yapılabilecek dönüşümler örneklerle anlatılıyor:

1. Eğer orijinal model  $Y = b_0 * e^{(b_1 * X)}$  şeklindeyse, doğrusal bir forma dönüştürmek için iki adımlı bir işlem yapılır:
  - Önce,  $Y$ 'nin doğal logaritması alınır:  $\ln(Y) = \ln(b_0) + b_1 * X$
  - Ardından,  $Z = \ln(Y)$  ve  $Z_2 = \ln(b_0)$  olarak tanımlanır. Bu durumda,  $Z = b_2 + b_1 * X$  lineer bir model elde edilir.
2. Eğer orijinal model  $Y = e^{(b_0)} * e^{(b_1 * X)}$  şeklindeyse, doğrusal bir forma dönüştürmek için şu adımlar izlenir:
  - Yine,  $Y$ 'nin doğal logaritması alınır:  $\ln(Y) = b_0 + b_1 * X$
  - Bu durumda,  $Z = \ln(Y)$  olarak tanımlanır ve bu,  $Z = b_0 + b_1 * X$  olarak lineer bir modele işaret eder.
3. Eğer orijinal model  $Y = (b_0 + b_1 * X)^2$  şeklindeyse, karekök alarak doğrusallaştırılabilir:
  - $\sqrt{Y} = b_0 + b_1 * X$
  - Burada,  $Z = \sqrt{Y}$  tanımlanır ve sonuç  $Z = b_0 + b_1 * X$  lineer bir modeldir.
4. Son örnek olarak, eğer orijinal model  $Y = \frac{1}{(b_0 + b_1 * X)}$  şeklindeyse, tersini alarak doğrusallaştırılır:
  - $\frac{1}{Y} = b_0 + b_1 * X$
  - Bu durumda,  $Z = \frac{1}{Y}$  olarak tanımlanır ve  $Z = b_0 + b_1 * X$  lineer bir model elde edilir.

Bu dönüşümler, istatistiksel analizlerde sıkça başvurulan yöntemlerdir ve doğrusal olmayan ilişkileri içeren verilerin lineer regresyon modellerinde kullanılabilmesini sağlar.





## Nonlinear regression (iterative estimation)

- Data =  $\{x_i, y_i\} \quad i=1..n$  ( $n$  = number of data points)
- $y = f(\beta, x)$
- $x = n \times d$  matrix
- $y = n \times 1$  matrix
- $r_i = y_i - f(\beta, x_i)$   $r$  = residuals ( $n \times 1$  matrix)
- $\beta$  = parameters to be optimized
- $E(\beta) = \sum (r_i)^2 \quad i=1..n$
- $\min_{\beta} E(\beta)$
- $dE(\beta)/d\beta = 0$  (optimize  $E$  according to  $\beta$ )

## Nonlinear regression (iterative estimation)

- $dE(\beta)/d\beta = 2 \cdot r \cdot dr/d\beta$
- $dr/d\beta = n \times d$  matrix =  $\begin{bmatrix} dr_1/d\beta_1 & dr_1/d\beta_2 & \dots & dr_1/d\beta_d \\ dr_2/d\beta_1 & dr_2/d\beta_2 & \dots & dr_2/d\beta_d \\ \dots & \dots & \dots & \dots \\ dr_n/d\beta_1 & dr_n/d\beta_2 & \dots & dr_n/d\beta_d \end{bmatrix}$
- $dr/d\beta$  is called Jacobian matrix ( $J$ )
- $\beta_{k+1} = \beta_k - \text{eps} \cdot dE(\beta)/d\beta$  (Gradient descent)
- $\beta_{k+1} = \beta_k - \text{eps} \cdot J^T \cdot r$  (Gradient descent)
- $(d \times 1) = (d \times 1) - \text{eps} \cdot (d \times n) \cdot (n \times 1)$



### ChatGPT

Gönderilen görselde doğrusal olmayan regresyon analizi ve bu analiz türünde parametre optimizasyonunun nasıl yapılacağı açıklanmaktadır. Bu süreç genellikle iteratif bir tahmin yöntemi kullanılarak gerçekleştirilir ve şu adımları içerir:

1. Veri noktaları  $\{x_i, y_i\}$  ile  $i = 1$ 'den  $n$ 'ye kadar olan bir dizi veri noktasını temsil eder, burada  $n$ , veri noktalarının sayısını ifade eder.
2.  $y$  yanıt vektörü fonksiyon  $f(\beta, x)$  ile hesaplanır,  $x$  bağımsız değişken matrisidir ve  $y$  bağımlı değişken vektörüdür.
3.  $r_i = y_i - f(\beta, x_i)$  denklemi ile her bir veri noktası için rezidüaller (artıklar) hesaplanır. Burada  $r$  rezidüallerin vektörüdür.
4.  $\beta$  parametreleri, hata fonksiyonu  $E(\beta)$ 'yi minimize edecek şekilde optimize edilmelidir. Hata fonksiyonu  $E(\beta) = \sum (r_i)^2$  olarak tanımlanır.
5. Optimal  $\beta$  değerlerini bulmak için, hata fonksiyonunun türevi  $\frac{dE(\beta)}{d\beta}$  sıfıra eşitlenir ve bu şekilde  $\beta$  parametreleri için en iyi değerler bulunur.

Görselin alt kısmında ise bu optimizasyon işleminin matematiksel ifadesi verilmiştir:

- $\frac{dE(\beta)}{d\beta} = 2 * r^T * dr/d\beta$  denklemi ile hata fonksiyonunun  $\beta$ 'ya göre türevi hesaplanır.
- $dr/d\beta$  ifadesi,  $\beta$ 'nın türevlerinin matrisidir ve Jakobyan matrisi olarak adlandırılır.
- Gradient descent (gradyan inişi) yöntemi ile bir sonraki  $\beta$  tahmini  $\beta_{k+1} = \beta_k - \epsilon * \frac{dE(\beta)}{d\beta}$  olarak güncellenir.
- Ayrıca, bir alternatif gradyan iniş formülü  $\beta_{k+1} = \beta_k - \epsilon * J^T * r$  olarak verilmiştir. Burada  $J^T$  Jakobyan matrisinin transpozesidir.
- $d(k+1) = d(k) - \epsilon * d(k) * n * 1$  ifadesi ile iteratif bir şekilde hata düzeltmesi yapılır.

Bu süreç genellikle bir dizi iterasyonla sürdürülür ve her iterasyonda  $\beta$  parametreleri daha uygun bir hale getirilir. Bu yöntem genellikle makine öğrenmesi ve istatistiksel modelleme alanlarında karmaşık modellerin parametrelerini tahmin etmek için kullanılır.

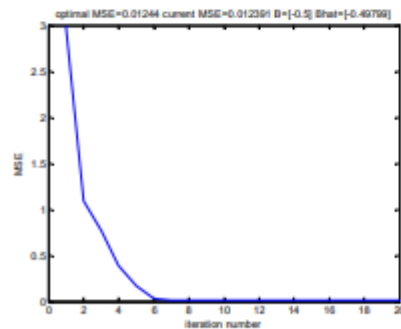
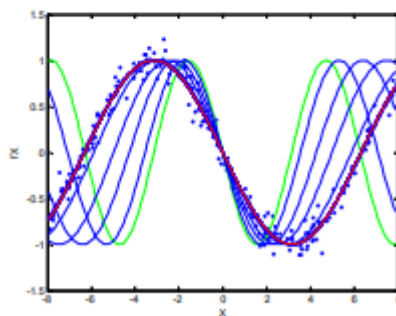


## Nonlinear regression (iterative estimation)

- $\beta_{k+1} = \beta_k - \text{eps} * dE(\beta)/d\beta$  (Gradient descent)
- $\beta_{k+1} = \beta_k - (dE(\beta)/d\beta) / (ddE(\beta)/dd\beta)$  (Newton Raphson)
- $ddE(\beta)/dd\beta \approx J^T * J$
- $\beta_{k+1} = \beta_k - (J^T * r) / (J^T * J)$
- $\beta_{k+1} = \beta_k - \text{inv}(J^T * J) * (J^T * r)$
- $\text{pinv}(J) = \text{inv}(J^T * J) * J^T$
- $\beta_{k+1} = \beta_k - \text{pinv}(J) * r$  (Newton Raphson)
- $\beta_{k+1} = \beta_k - \text{eps} * J^T * r$  (Gradient descent)

Blue points: data  
Green line: initial guess  
Blue lines: iterative guesses  
Red line: last guess

$$Y = \sin(b_1 * X)$$



Gradient descent

Eps=0.0005

Real b=-0.5

initial b=-1

Non\_lin\_reg0.m



## ChatGPT

Görselde doğrusal olmayan regresyon için iteratif tahmin yöntemleri ve bu yöntemlerin formülleri gösteriliyor. Yöntemlerden ilki gradyan inişi (Gradient Descent), ikincisi ise Newton-Raphson yöntemidir.

Gradyan İnişi yöntemi, bir fonksiyonun minimum noktasına iteratif olarak yaklaşmayı amaçlar. Bu yöntemde, her adımda parametreler ( $\beta$ ), hata fonksiyonunun gradyanının negatif yönünde bir adım atarak güncellenir:

$$\beta_{k+1} = \beta_k - \epsilon * \frac{dE(\beta)}{d\beta}$$

Burada  $\epsilon$ , öğrenme hızıdır ve  $\frac{dE(\beta)}{d\beta}$ , hata fonksiyonunun  $\beta$  parametresine göre türevidir.

Newton-Raphson yöntemi ise, bir fonksiyonun köklerini bulmak için kullanılan daha hızlı ama daha karmaşık bir yöntemdir. Bu yöntemde parametreler, hata fonksiyonunun ikinci türevine bağlı olarak güncellenir:

$$\beta_{k+1} = \beta_k - \frac{(dE(\beta)/d\beta)}{(d^2E(\beta)/d\beta^2)}$$

Ayrıca, hata fonksiyonunun Jakobyan matrisi  $J$  ile ifadesi verilmiş ve  $\beta$  parametrelerinin güncellenmesi için farklı formüller sunulmuştur:

$$dE(\beta)/d\beta = J^T * r$$

$$\beta_{k+1} = \beta_k - \epsilon * J^T * r$$

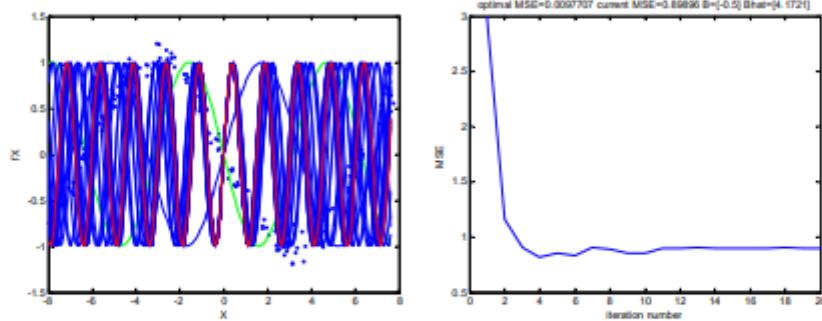
$$\beta_{k+1} = \beta_k - inv(J^T * J) * (J^T * r)$$

Burada  $inv()$  işlemi matrisin tersini almayı ifade eder.  $pinv(J)$  ise  $J$  matrisinin Moore-Penrose pseudoinversini belirtir.

Son olarak, görselin alt kısmında bir regresyon analizi örneği gösterilmiş. Model  $Y = \sin(b1 * X)$  olarak belirtilmiş ve bir dizi mavi nokta (veri), başlangıç tahminleri (yeşil çizgiler), iteratif tahminler (mavi çizgiler) ve son tahmin (kırmızı çizgi) şeklinde görselleştirilmiştir. Grafikte, belirli bir  $\epsilon$  değeri (öğrenme hızı) ve gerçek  $b$  değeri belirtilerek gradyan inişi yöntemi ile bu gerçek değere nasıl yaklaşıldığı gösteriliyor. Bu tür görselleştirmeler, modelin nasıl iyileştirildiğini ve iteratif tahmin yönteminin verileri nasıl modele uygun hale getirdiğini anlamak için kullanılır.



$$Y = \sin(b_1 * X)$$



Gradient descent

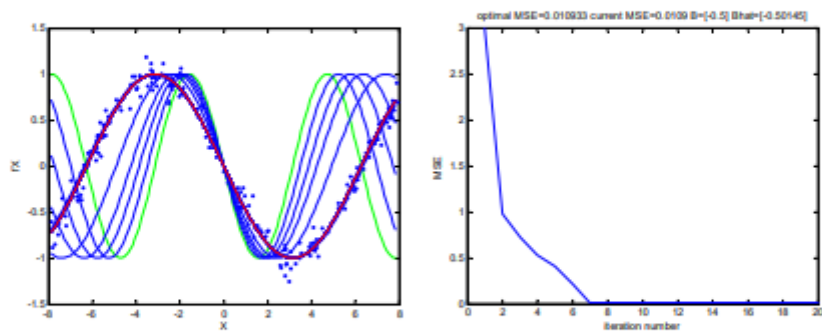
Eps=0.01

Real b=-0.5

initial b= -1

Non\_lin\_reg0.m

$$Y = \sin(b_1 * X)$$



Newton raphson

Real b=-0.5

initial b=-1

Non\_lin\_reg0.m



## ChatGPT

Görselde iki farklı optimizasyon tekniği olan Gradyan İnişi ve Newton-Raphson yöntemlerinin regresyon analizinde nasıl kullanıldığını gösteren iki grafik bulunuyor. Her iki teknik de  $Y = \sin(b * X)$  modelini kullanarak parametre tahmininde bulunur.

Üstteki grafikte Gradyan İnişi yöntemi ile yapılan optimizasyonun sonuçları gösteriliyor. Gradyan İnişi kullanılarak, öğrenme hızı ( $\epsilon$ ) 0.01 olarak ayarlanmış ve gerçek  $b$  değeri -0.5 olarak belirtilmiş. Sol taraftaki grafik, tahmin edilen  $b$  değerlerinin iterasyonlar boyunca nasıl değiştiğini gösteriyor ve görüldüğü üzere, tahminler gerçek değere yaklaşmakta zorlanıyor. Sağ taraftaki grafik ise hata değerlerinin iterasyon sayısı ile nasıl azaldığını gösteriyor.

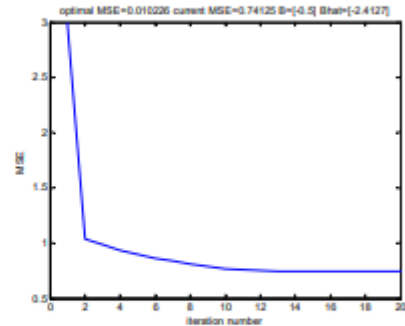
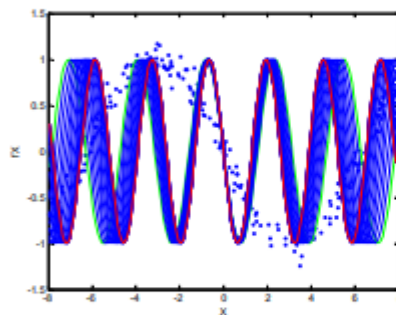
Alt grafikte ise Newton-Raphson yöntemi ile yapılan optimizasyon sonuçları yer alıyor. Burada başlangıçta  $b$  değeri -1 olarak ayarlanmış ve yine gerçek  $b$  değeri -0.5. Sol taraftaki grafikte, Newton-Raphson yöntemi ile yapılan tahminlerin her iterasyonda daha da iyileştiği ve gerçek değere hızla yaklaştığı görülüyor. Sağ taraftaki grafik ise bu yöntemle hata değerlerinin hızla düştüğünü gösteriyor.

Her iki yöntem de parametre tahmininde farklı yaklaşımlar sunar ve bu örnekler, her birinin iteratif olarak nasıl daha iyi tahminlere ulaştığını görselleştirir. Gradyan İnişi genellikle daha basittir ve geniş ölekte uygulanabilirken, Newton-Raphson yöntemi daha az iterasyonda daha hızlı sonuç verebilir ancak her durumda uygulanabilir olmayabilir.





$$Y = \sin(b_1 * X)$$



Newton raphson

Real b=-0.5

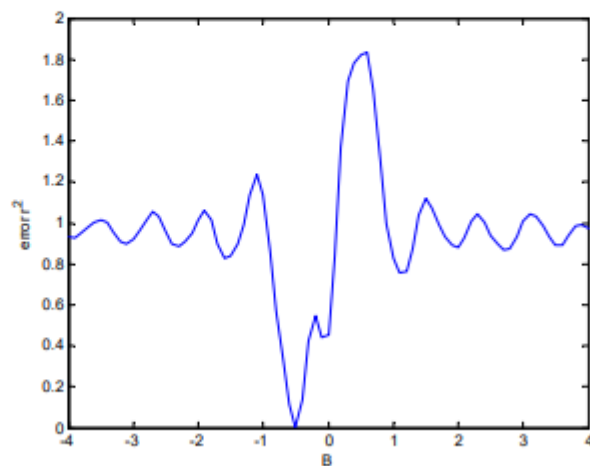
initial b=-2

Non\_lin\_reg0.m

$$Y = \sin(b_1 * X)$$

b1 vs. error

- B1=-0.5



Non\_lin\_reg1.m



### ChatGPT

Görselde, Newton-Raphson yöntemi kullanılarak yapılan bir doğrusal olmayan regresyon analizinin sonuçları yer alıyor. Bu analiz,  $Y = \sin(b1 * X)$  modeli için yapılmıştır.

Üstteki grafiklerde, gerçek  $b$  değeri -0.5 iken başlangıçta  $b$  değeri -2 olarak ayarlanmış. Sol taraftaki grafik, tahmin edilen  $b$  değerlerinin her iterasyonda nasıl değiştiğini gösteriyor. Görülebileceği üzere, değerler dalgalanarak gerçek değere yaklaşıyor. Sağ taraftaki grafik ise hata değerlerinin (yani tahmin edilen değerler ile gerçek değer arasındaki farkın) iterasyon sayısı ile nasıl azaldığını gösteriyor ve hata değerinin zamanla düştüğü görülüyor.

Altta yer alan grafik, farklı  $b1$  değerleri için hata değerlerini gösteriyor. Burada,  $b1$  değeri -0.5 olarak belirtilmiş ve  $b1$  değerinin hata üzerindeki etkisi görselleştirilmiş. Grafikte,  $b1$  değerinin değişiminin hata üzerinde yarattığı dalgalanmalar görülebiliyor.

Bu grafikler, optimizasyon sürecinin nasıl işlediğini ve farklı başlangıç parametre değerlerinin sonuçları nasıl etkileyebileceğini gösteriyor. Newton-Raphson yöntemi, tahmin edilen parametre değerlerini gerçek değerlere iteratif bir şekilde yaklaştırmak için kullanılır. Bu süreç, veri uyumu ve model kalitesini artırma amacı taşır.



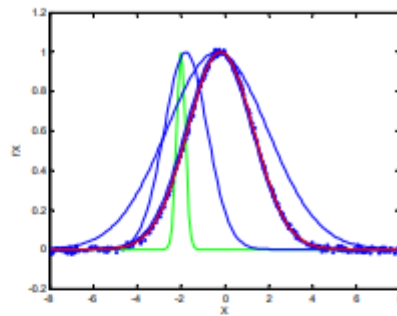
Blue points: data

Green line: initial guess

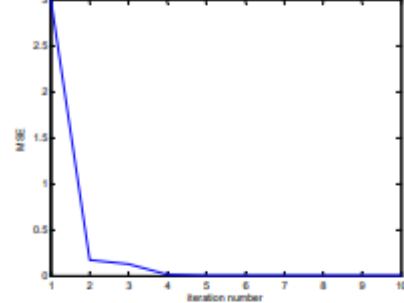
Blue lines: iterative guesses

Red line: last guess

$$Y = \exp(-(X-B1)^2/(2*B2^2))$$



optimal MSE=1.5114e-005 current MSE=0.00016955 B=[-0.2 1.5] Bhat=[-0.2082 -1.4744]



Gradient descent

Eps=0.1

Real B=[-0.2 ; 1.5]

initial B= [-2; -0.2]

Non\_lin\_reg2.m

Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG DEPT.

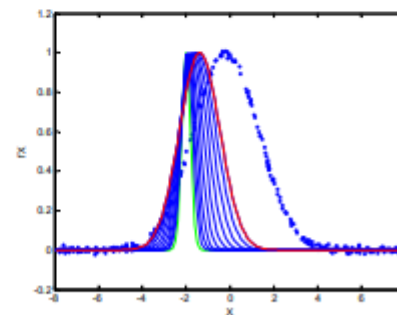
Blue points: data

Green line: initial guess

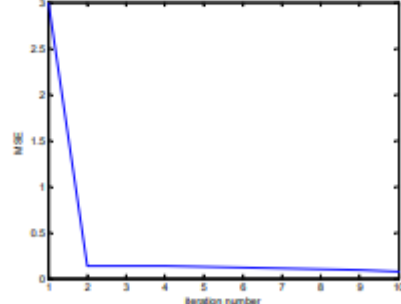
Blue lines: iterative guesses

Red line: last guess

$$Y = \exp(-(X-B1)^2/(2*B2^2))$$



optimal MSE=3.0804e-007 current MSE=0.076773 B=[-0.2 1.5] Bhat=[-1.3777 -0.92512]



Gradient descent

Eps=0.01

Real B=[-0.2 ; 1.5]

initial B= [-2; -0.2]

Non\_lin\_reg2.m

Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG DEPT.

Windows 7



### ChatGPT

Görselde, doğrusal olmayan bir regresyon modeli için Gradyan İnişi yöntemi ile parametre optimizasyonu yapılırken kullanılan iterasyon sürecinin iki farklı örneği yer almaktadır. Her iki örnekte de kullanılan model  $Y = \exp(-(X - B1)^2 / (2 * B2^2))$  şeklindedir ve bu, bir Gauss (veya normal) dağılım fonksiyonunu ifade eder.

İlk örnekte, Gradyan İnişi için öğrenme hızı ( $\epsilon$ ) olarak 0.1 kullanılmış. Gerçek parametre değerleri  $B = [-0.2, 1.5]$  iken, başlangıç tahminleri  $B = [-2, -0.2]$  olarak ayarlanmış. Sol taraftaki grafik, iterasyonlar boyunca tahmin edilen parametre değerlerinin nasıl değiştiğini gösteriyor ve bu değerlerin zamanla gerçek değerlere yaklaştığını gösteren mavi çizgilerle belirtilmiş. Sağ taraftaki grafik ise hata değerlerinin (model tahminlerinin gerçek değerlerden ne kadar sapma gösterdiğini gösteren) iterasyon sayısına bağlı olarak nasıl azaldığını gösteriyor.

İkinci örnekte, öğrenme hızı daha düşük bir değer olan 0.01 olarak belirlenmiş ve aynı gerçek ve başlangıç parametre değerleri kullanılmış. Burada da sol taraftaki grafik, tahmin edilen parametrelerin her iterasyonda nasıl iyileştiğini gösteriyor ve sağ taraftaki grafik hata değerlerinin azalışını gösteriyor.

Her iki örnekte de, iterasyonların ilerlemesiyle birlikte tahminlerin gerçek değerlere yaklaştığı ve hata oranlarının düştüğü görülüyor. Bu tür görselleştirmeler, optimizasyon yöntemlerinin performansını değerlendirmek ve model parametrelerinin nasıl güncellendiğini anlamak için kullanılır.

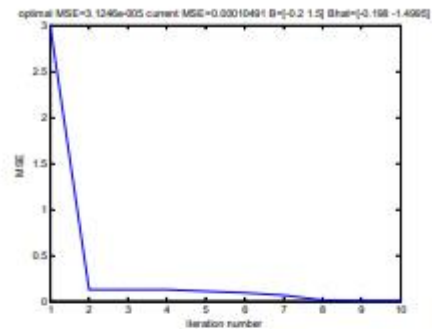
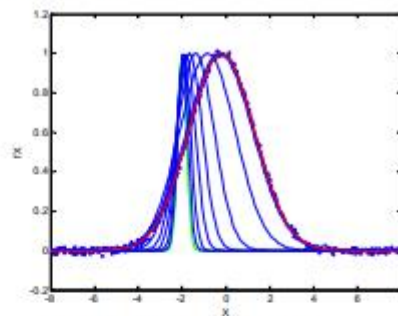
Blue points: data

Green line: initial guess

Blue lines: iterative guesses

Red line: last guess

$$Y = \exp(-(X-B1)^2/(2*B2^2))$$



Newton raphson

Real B=[-0.2 ; 1.5]

initial B= [-2; -0.2]

Non\_lin\_reg2.m

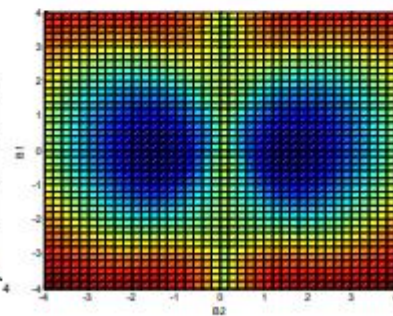
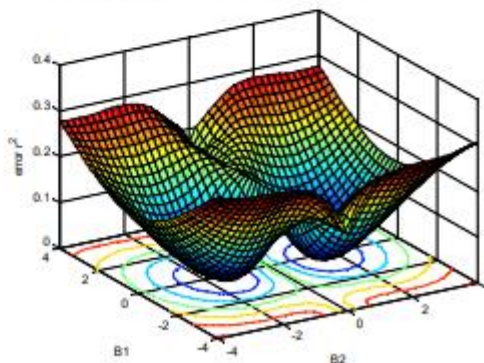
Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG DEPT.

$$Y = \exp(-(X-B1)^2/(2*B2^2))$$

b vs. error

- N=200; B=[-0.2 ; 1.5]



Why symmetric?

Non lin reg3.m

Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG DEPT.

Windows'u Etkir



## ChatGPT

Bu görsel, bir dizi optimizasyon tekniklerini ve bunların sonuçlarını gösteren grafikler içeren bir ders notundan alıntı gibi görünüyor. Ders notları, "Optimizasyon Teknikleri" başlığını taşıyan Mehmet Fatih Amasyalı'ya ait.

Üstteki iki grafik, verileri (mavi noktalar) ve bu verilere uyan en iyi eğriyi bulmak için kullanılan iteratif bir yaklaşımın (Newton-Raphson yöntemi) adımlarını gösteriyor. Yatay eksenler değişkeni ( $X$ ), dikey eksenler ise  $Y$ 'nin değerini temsil ediyor. Yeşil çizgi ilk tahminleri, mavi çizgiler iteratif tahminleri ve kırmızı çizgi son tahmini gösteriyor. İkinci grafik, hata miktarının optimizasyon iterasyonları boyunca nasıl değiştiğini gösteriyor. Hata azaldıkça, model verilere daha iyi uymaya başlıyor.

Altteki grafiklerden biri,  $B_1$  ve  $B_2$  parametrelerine göre hatanın nasıl değiştiğini 3 boyutlu bir yüzey olarak gösteriyor. Diğer grafik ise bu hatanın  $B_1$  ve  $B_2$  parametreleri için 2 boyutlu bir ısı haritası olarak gösteriliyor ve simetrik bir yapı sergilediği soruluyor. " $N=200$ ,  $B=[-0.2; 1.5]$ " ifadesi, deneyin 200 veri noktası içerdiğini ve  $B$  parametrelerinin aralığının  $-0.2$  ile  $1.5$  arasında olduğunu gösteriyor.

Bu tür grafikler genellikle verilerden belirli bir modelin parametrelerini tahmin etmek için kullanılır ve bu örnekte, bir optimizasyon tekniği kullanılarak en uygun parametre değerleri bulunmaya çalışılmış.



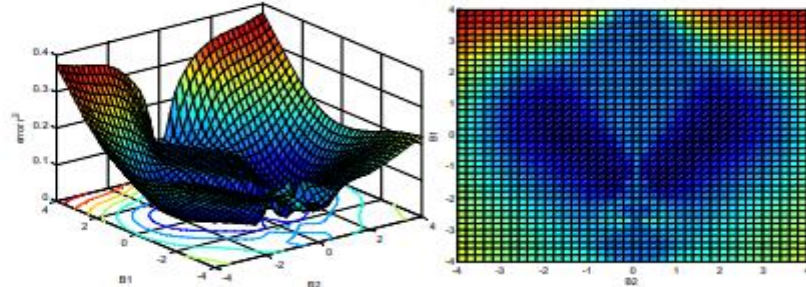


## Sample size effect

$$Y = \exp(-(X-B1)^2/(2*B2^2))$$

b vs. error

- $N=10$ ;  $B=[-0.2 ; 1.5]$



It is more difficult. Why?

Non lin reg3.m

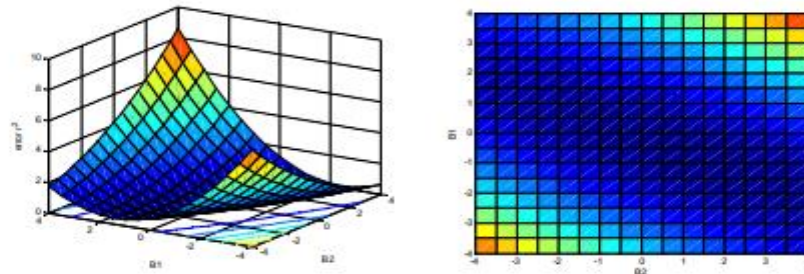
Mehmet Fatih AMAŞYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG DEPT.

$$Y = b1*x + b2*x^3 \text{ (linear according to B's)}$$

b vs. error

- $N=10$ ;  $B=[-1 ; 1]$



It is quadratic.

Non lin reg4.m

Mehmet Fatih AMAŞYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG DEPT.



## ChatGPT

Bu görseller, farklı örneklem boyutları ve parametre aralıkları ile gerçekleştirilen iki farklı non-lineer regresyon analizini gösteren ders notlarından alınmış. Her iki durumda da, amacımız verilerimize en iyi uyan model parametrelerini bulmak.

İlk grafikler seti, "Örneklem büyüklüğü etkisi" başlığı altında yer alıyor. Burada  $N=10$ , yani 10 veri noktası kullanılmış ve B parametreleri için aralık  $[-0.2; 1.5]$  olarak verilmiş. Fonksiyon,  $Y = \exp\left(-\frac{(X-B_1)^2}{2*B_2^2}\right)$  şeklinde tanımlanmış. 3 boyutlu grafikte B1 ve B2 parametrelerine karşı hatanın nasıl değiştiğini görebiliyoruz. 2 boyutlu renkli grafik ise bu hatayı B1 ve B2'nin her bir kombinasyonu için bir ısı haritası olarak gösteriyor ve bu durumun neden daha zor olduğunu sorguluyor.

İkinci grafikler setinde, fonksiyon doğrusal olarak B parametrelerine bağlı olarak tanımlanmış:  $Y = b_1 * X + b_2 * X^3$ . Burada da  $N=10$ , yani yine 10 veri noktası kullanılmış ve B parametreleri için aralık  $[-1; 1]$  olarak belirlenmiş. Bu durumda, 3 boyutlu grafik ve 2 boyutlu renkli grafik, hatanın B1 ve B2'ye bağlı olarak nasıl değiştiğini gösteriyor ve hatanın karesel olduğu belirtiliyor.

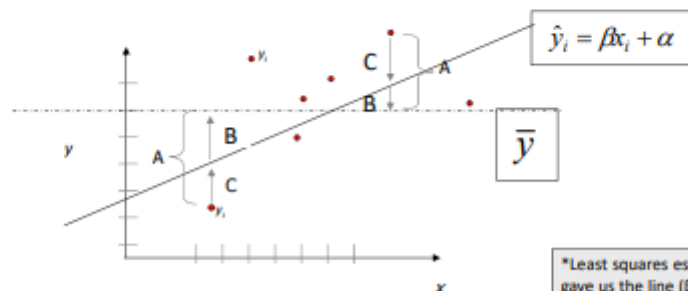
Her iki durumda da, grafikler parametre tahminlerinin nasıl yapıldığını ve bu tahminlerin hatası ile nasıl ilişkilendirildiğini göstermek için kullanılıyor. Grafikler ve kod isimleri (non\_lin\_reg3.m ve non\_lin\_reg4.m), MATLAB gibi bir programlama dilinde yazılmış non-lineer regresyon scriptlerini işaret ediyor olabilir.



## Modeling Interactions

- Statistical Interaction: When the effect of one predictor (on the response) depends on the level of other predictors.
- Can be modeled (and thus tested) with cross-product terms (case of 2 predictors):
- $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$

## Regression Picture



$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (\hat{y}_i - y_i)^2$$

$A^2$

$SS_{total}$

Total squared distance of observations from naive mean of y  
Total variation

$B^2$

$SS_{reg}$

Distance from regression line to naive mean of y  
Variability due to x (regression)

$C^2$

$SS_{residual}$

Variance around the regression line  
Additional variability not explained by x—what least squares method aims to minimize

$$R^2 = SS_{reg} / SS_{total}$$



## ChatGPT

Bu görsel, istatistiksel etkileşimlerin modellenmesi ve regresyon analizi ile ilgili konuları içeren bir ders notunun parçası.

**İstatistiksel Etkileşim:** Bir yanıt değişkeni üzerindeki bir tahmincinin (prediktor) etkisinin, diğer tahmincilerin seviyesine bağlı olarak değiştiği durum. Örneğin, bir ilacın etkisinin hastanın yaşı gibi başka bir faktöre bağlı olması gibi düşünülebilir.

Bu tür etkileşimler, çapraz çarpım terimleri (iki tahminci durumu) kullanılarak modellenenir ve test edilebilir. Formüldeki  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$  ifadesi, yanıt değişkeni  $Y$ 'nin, iki bağımsız değişken  $X_1$  ve  $X_2$  ve bu bağımsız değişkenlerin etkileşimi olan  $X_1 X_2$  teriminin lineer kombinasyonu olarak modellendiğini gösterir.

**Regresyon Grafiği:** En alttaki grafik ise regresyon analizinde kullanılan temel kavramları görselleştiriyor. Gerçek veri noktaları ( $y$ ) ve regresyon doğrusunun ( $\hat{y} = \beta_x + \alpha$ ) tahmin ettiği değerler arasındaki farkları ( $C$ ), yani karelerin toplamını ( $SS_{total}$ ), en aza indirgeyerek en iyi doğruyu bulmayı hedefleyen en küçük kareler yöntemini açıklıyor. Burada  $\hat{y}$ ,  $y$ 'nin tahmin edilen değerini,  $\beta_x$  eğimi ve  $\alpha$  ise  $y$  kesme noktasını ifade ediyor. Ayrıca,  $R^2$  değeri, regresyonun toplam varyansın ne kadarını açıkladığını gösterir ve 0 ile 1 arasında bir değer alır.  $R^2$  ne kadar yüksekse, model o kadar iyi bir uyum sağlamış demektir.

Bu görseldeki formüller ve açıklamalar, regresyon modelinin ne kadar iyi olduğunu değerlendirmek için kullanılır ve modelin veriye ne kadar iyi uyduğunu anlamak için istatistiksel testler yapmaya olanak tanır.