

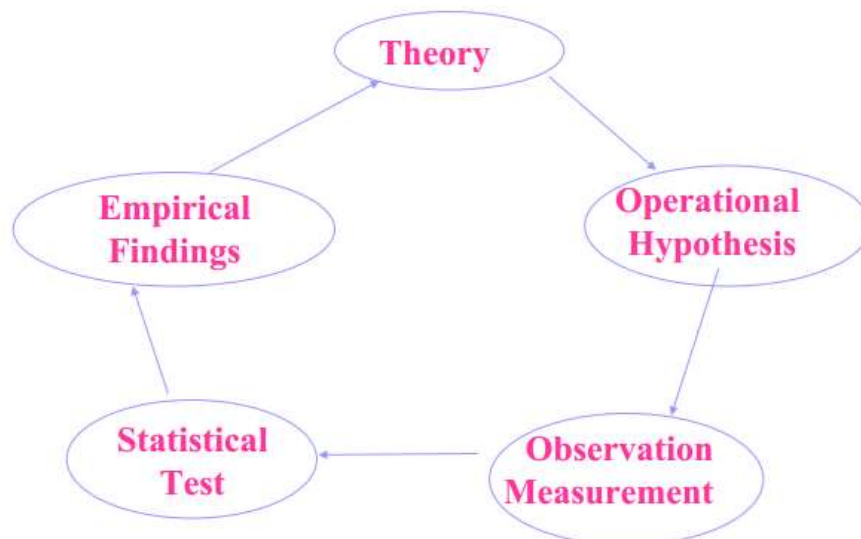
Data Modelling and Regression Techniques

M. Fatih Amasyalı

Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG DEPT.

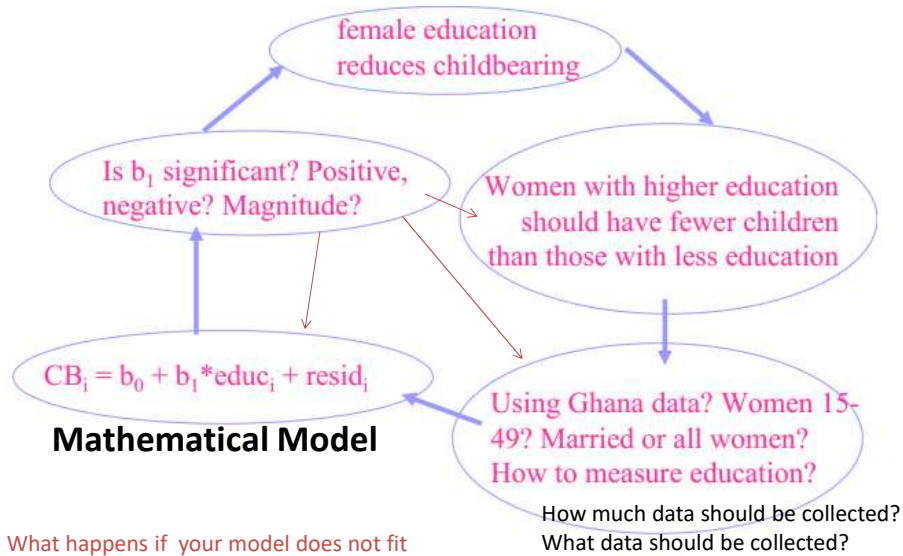
The traditional scientific approach



Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG DEPT.

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.

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

Specifying the deterministic component

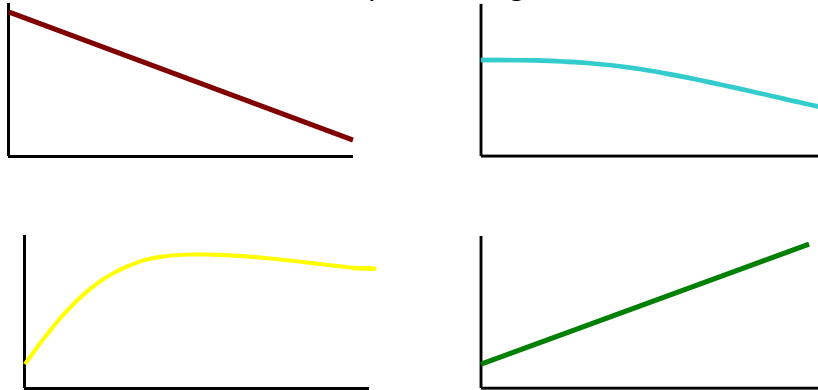
1. Define the dependent variable and independent variable(s)
2. Hypothesize Nature of Relationship
 - Expected Effects (i.e., Coefficients' Signs)
 - Functional Form (Linear or Non-Linear)
 - Interactions

Model Specification Is Based on Theory

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

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 $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_n X^n + \varepsilon$

ε is random error.

The word **linear** refers to the linearity of the parameters β_i .

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

Types of Regression Models

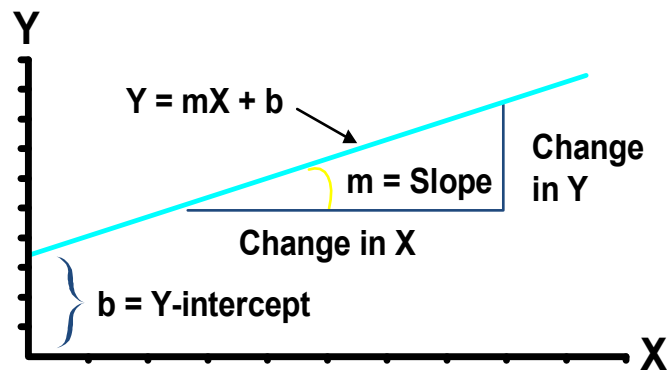
- If the parameters are linear related to the each other the model is linear. A non-linear first order model: $Y = (\beta_0 X) / (\beta_1 + X) + \epsilon$
- If X has d dimensions, a linear first order model: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_d X_d + \epsilon$

- A linear first order model $Y = \beta_0 + \beta_1 X + \epsilon$
- To get the model, we need to estimate the parameters β_0 and β_1
- Thus, the estimate of our model is

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

- Y_{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.

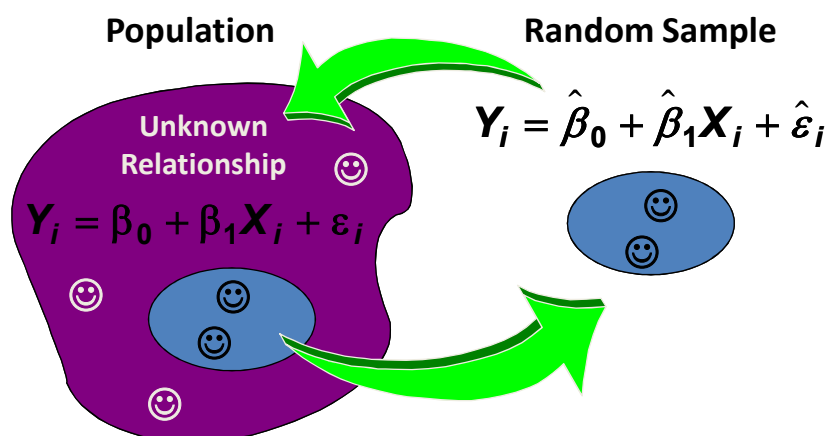
An Old Friend



Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

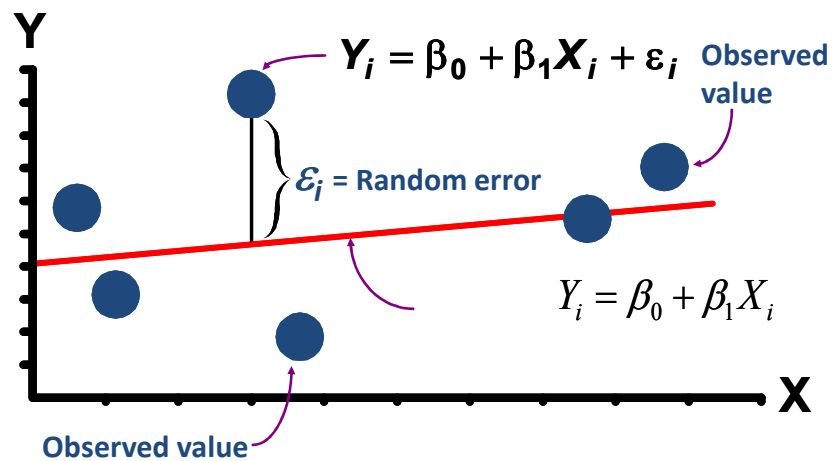
YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG. DEPT.

Population & Sample Regression Models



Mostly, you can not reach the all population.

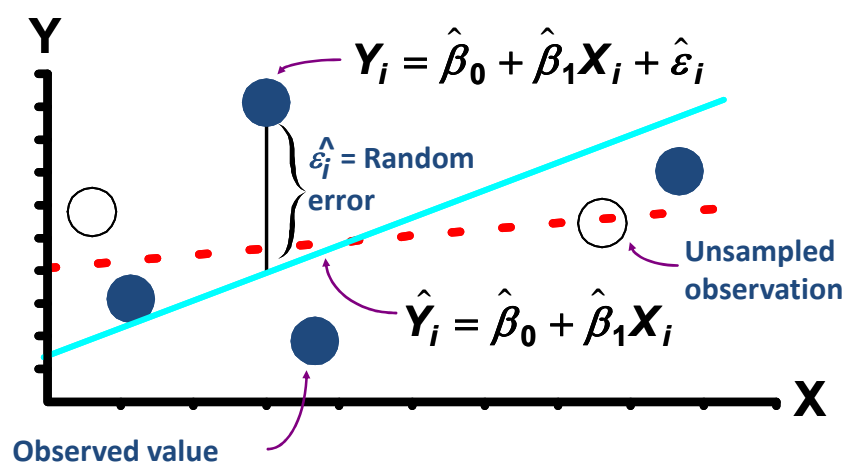
Population Linear Regression Model



Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG. DEPT.

Sample Linear Regression Model



Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG. DEPT.

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!

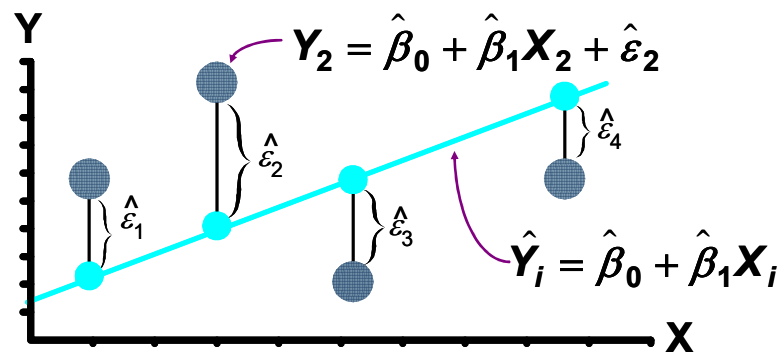
$$\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n \hat{\varepsilon}_i^2$$

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

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

Least Squares Graphically

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



Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG. DEPT.

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

Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG. DEPT.

Assume that our model is $Y=\beta$

- How can we estimate β using LS?
- Least Squares Minimize squared error

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \beta)^2$$

$$\frac{\partial \sum_{i=1}^n (y_i - \beta)^2}{\partial \beta} = 0$$

$$-2 \sum_{i=1}^n (y_i - \beta) = 0$$

$$\sum_{i=1}^n y_i - \sum_{i=1}^n \beta = 0$$

$$\sum y = n\beta$$

$$\beta = \frac{1}{n} \sum y$$

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

Example-1

- $Y_{n \times 1} = X_{n \times (d+1)} \beta_{(d+1) \times 1} + \varepsilon_{n \times 1}$

- $n=4, d=1$

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

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

Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG. DEPT.

Example-2

- $Y_{n \times 1} = X_{n \times (d+1)} \beta_{(d+1) \times 1} + \varepsilon_{n \times 1}$

- $n=4, d=2$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ 1 & x_{31} & x_{32} \\ 1 & x_{41} & x_{42} \end{bmatrix} * \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

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

Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG. DEPT.

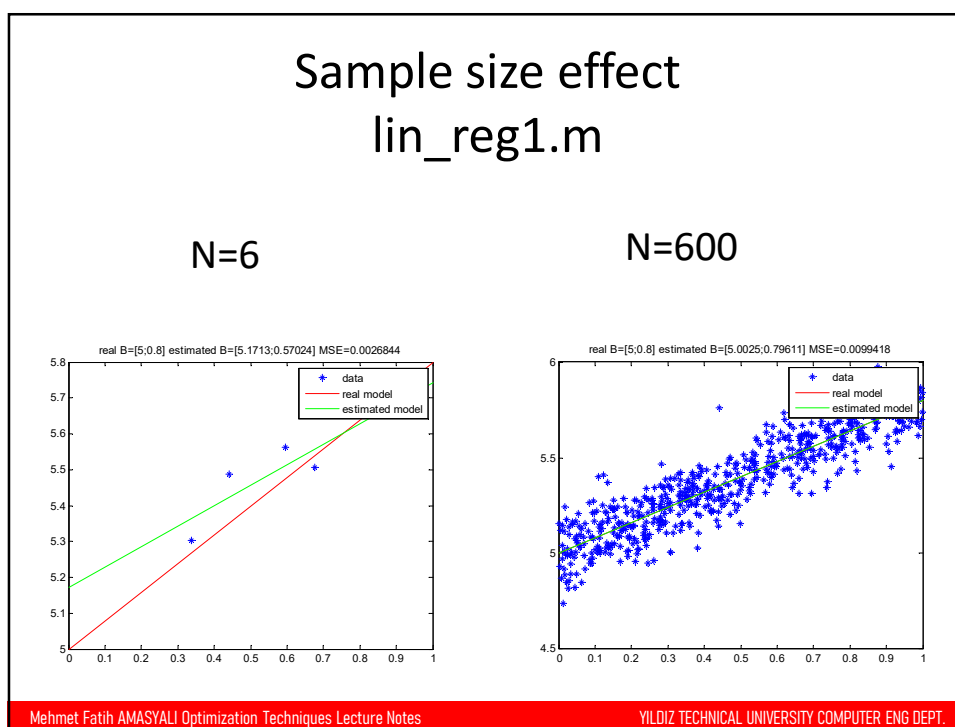
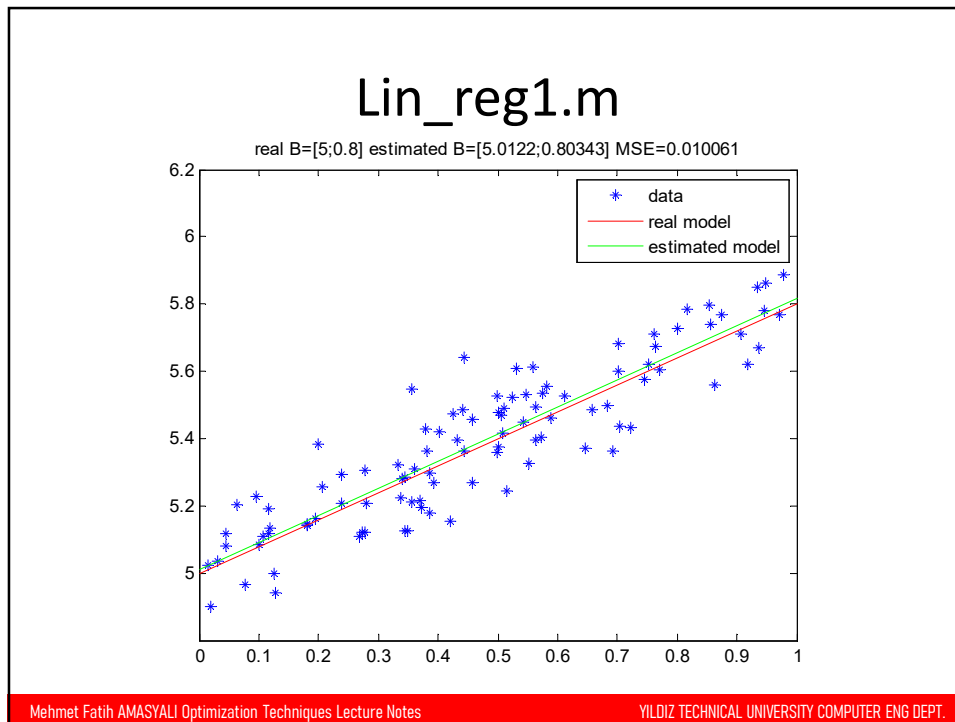
Example-3

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

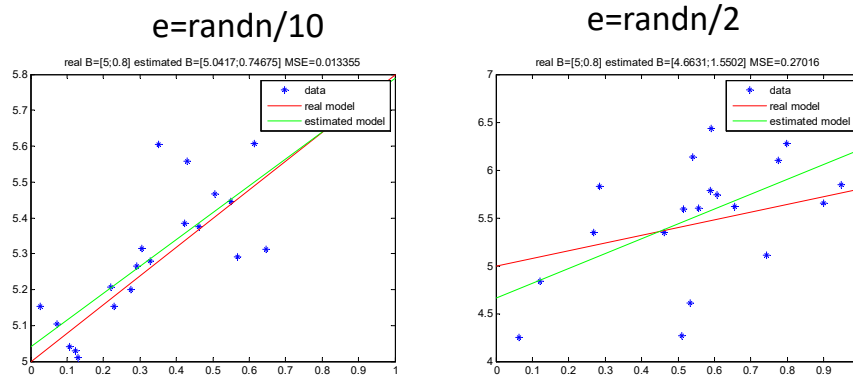
- $n=4, d=1, \text{order}=2$

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

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



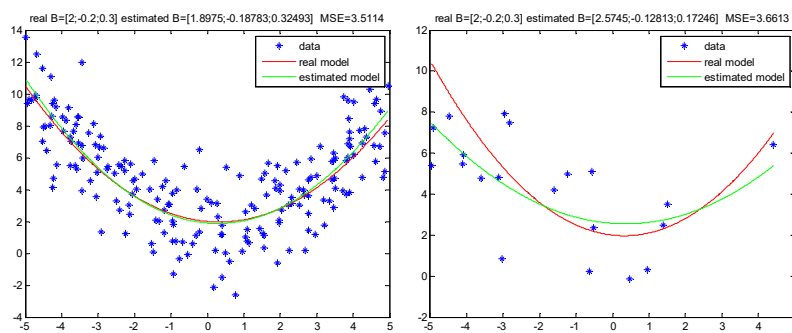
Error rate effect lin_reg1.m



Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG DEPT.

$Y=b_0+b_1*x+b_2*x^2$ Lin_reg2.m

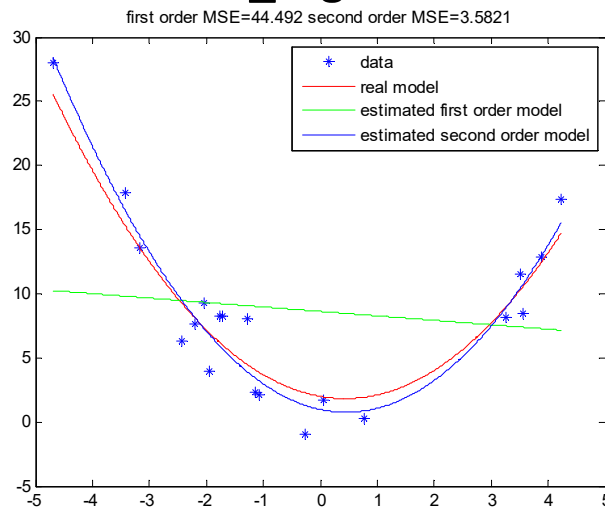


Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG DEPT.

Real 2nd order, estimated 1st and 2nd order

Lin_reg3.m

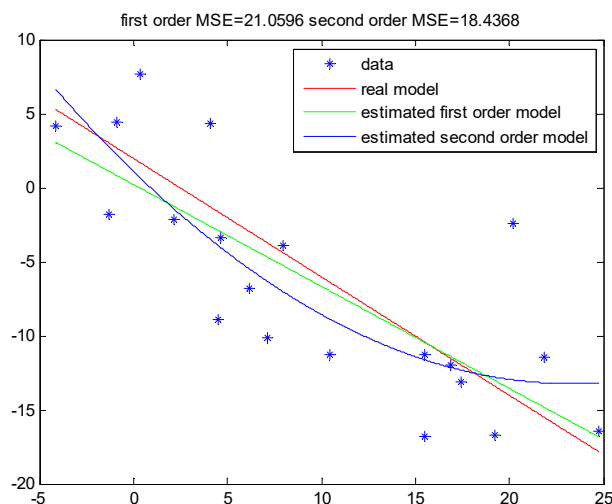


Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG. DEPT.

Real 1st order, estimated 1st and 2nd order

Lin_reg4.m



← What happens in these areas? →

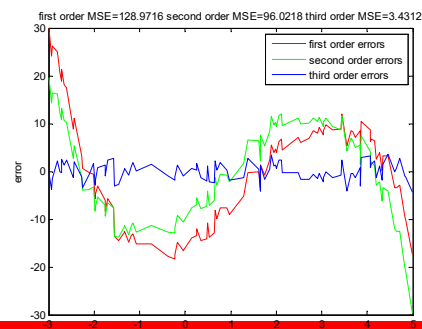
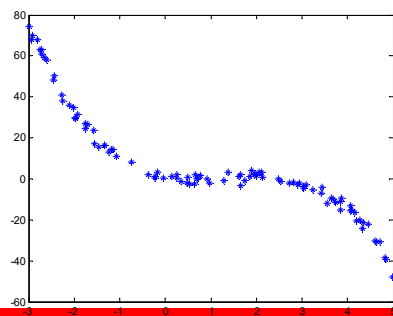
Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG. DEPT.


```
% 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 ]'
```

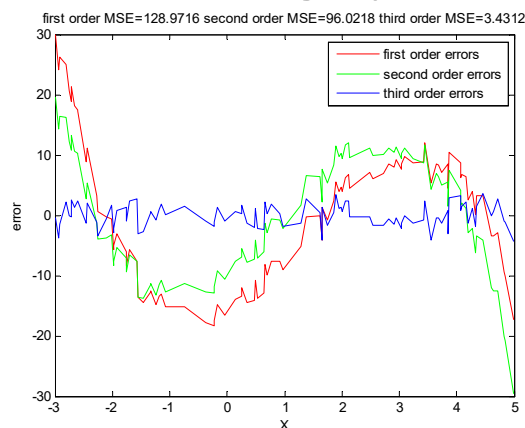
Lin_reg5.m



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.

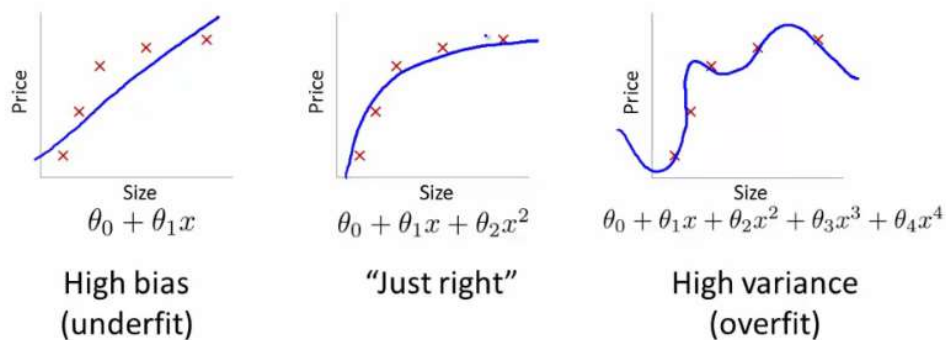
If the model degree is sufficient,
the errors have the same variances along X.

Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG DEPT.

Overfitting

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

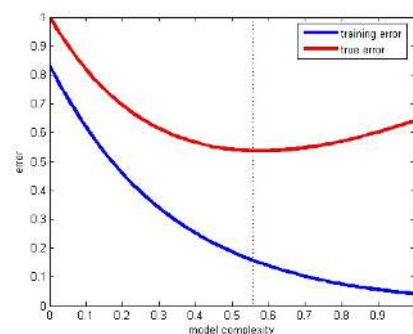


Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

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.

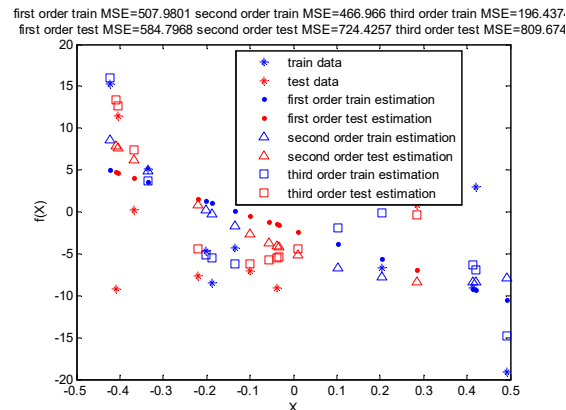


Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG. DEPT.

Model Validation

- Training and test sets are separated data samples.
- Lin_reg6.m



Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG. DEPT.

$Y=X*\beta$ Linear System Construction

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

Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG. DEPT.

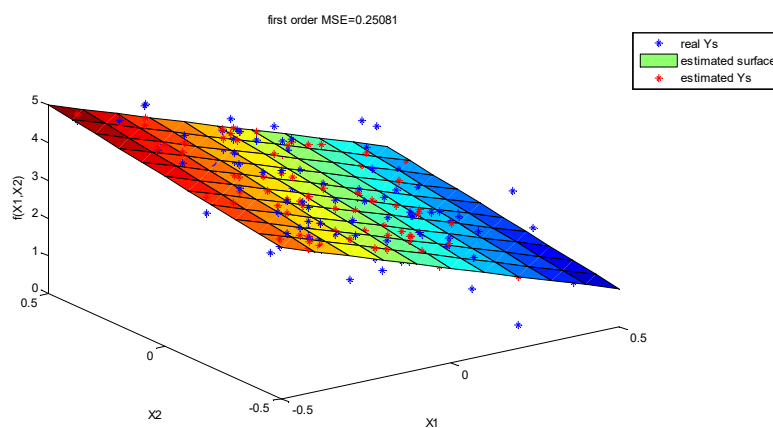
$Y=X*\beta$ Linear System Construction

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

Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG. DEPT.

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



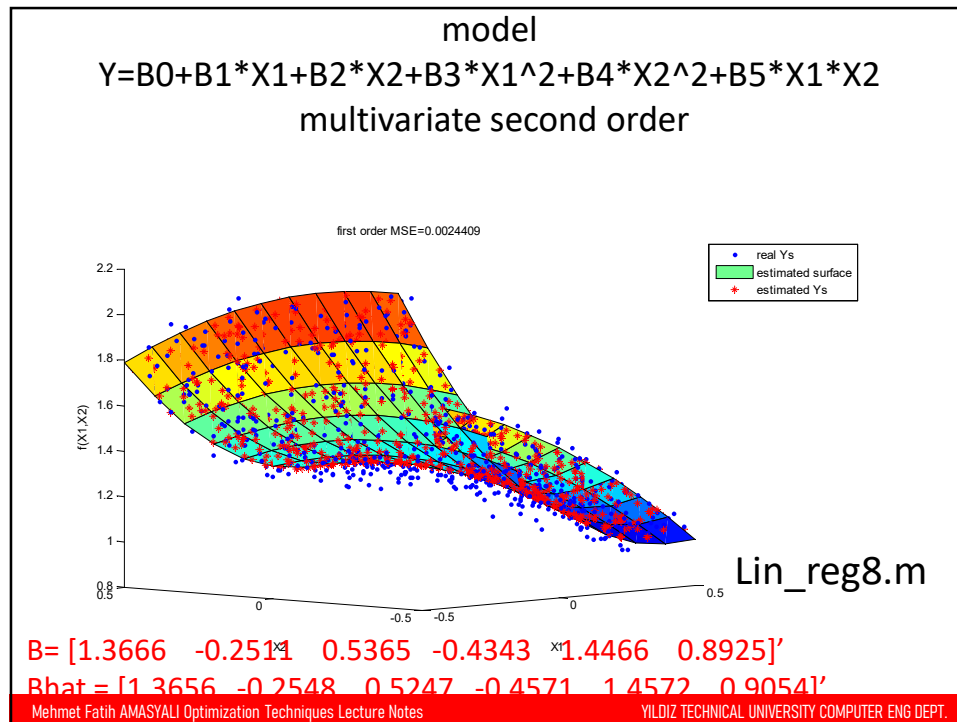
$$B = [3 \quad -3 \quad 1]'$$

$$\text{Bhat} = [3.0216 \quad -3.0439 \quad 1.0387]'$$

Lin_reg7.m

Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG. DEPT.



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)
 - $Y=\sin(b_1*X)$
 - $Y=b_1+\sin(b_2*X)$
 - $Y=\exp(b_1*X)$
 - $Y=(b_1*X)/(b_2+X)$
 - $Y=b_1*(\exp -((X-b_2)^2/(b_3^2)))$

Non-linearity

- A parameter β of the function f appears nonlinearly if the derivative $\partial f / \partial \beta$ is a function of β .
- The model $M(\beta, x)$ is nonlinear if at least one of the parameters in β appear nonlinearly.
- $f(x) = \beta \sin(x)$, $\partial f / \partial \beta = \sin(x)$, which is independent of β , 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 β , 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 β_1 , $\partial f / \partial \beta_2 = \cos(x)$, which is independent of β_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 β_1 , but $\partial f / \partial \beta_2 = -x \sin(\beta_2 x)$, which is dependent of β_2 , so the model $M(\beta, x)$ is non-linear.

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

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)

Nonlinear regression (iterative estimation)

- Data = $\{x_i, y_i\} \ i=1..n$ (n= number of data points)
- $y=f(\beta, x)$
- $x = n*d$ matrix
- $y = n*1$ matrix
- $r_i = y_i - f(\beta, x_i)$ r = residuals ($n*1$ matrix)
- β = 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 β)

Nonlinear regression (iterative estimation)

- $dE(\beta)/d\beta = 2*r*dr/d\beta$
- $dr/d\beta = n*d$ 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)
- $\beta_{k+1} = \beta_k - \text{eps} * dE(\beta)/d\beta$ (Gradient descent)
- $\beta_{k+1} = \beta_k - \text{eps} * J^T * r$ (Gradient descent)
- $(d*1) = (d*1) - \text{eps} * (d*n) * (n*1)$

Nonlinear regression (iterative estimation)

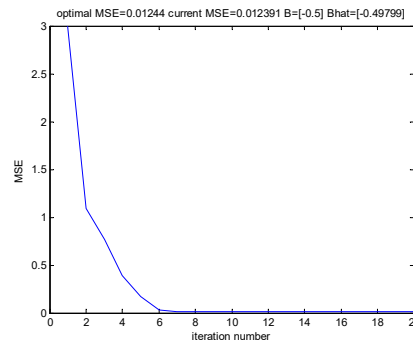
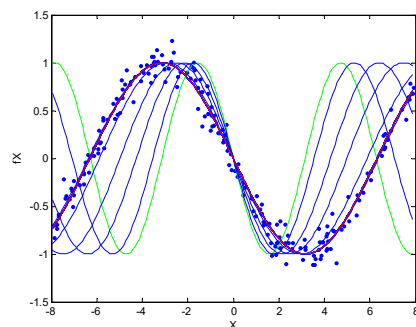
- $\beta_{k+1} = \beta_k - \text{eps} * dE(\beta)/d\beta$ (Gradient descent)
- $\beta_{k+1} = \beta_k - (dE(\beta)/d\beta) / (ddE(\beta)/dd\beta)$ (Newton Raphson)
- $ddE(\beta)/dd\beta \approx J^T * J$
- $\beta_{k+1} = \beta_k - (J^T * r) / (J^T * J)$
- $\beta_{k+1} = \beta_k - \text{inv}(J^T * J) *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

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

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



Gradient descent

Eps=0.0005

Real b=-0.5

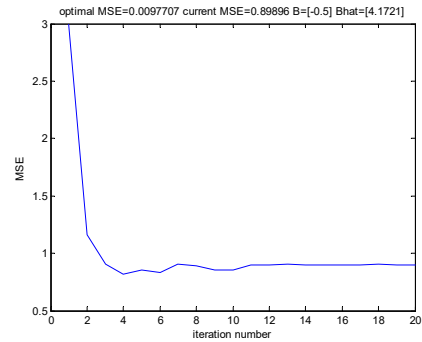
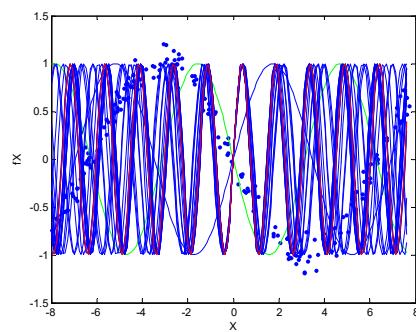
initial b=-1

Non_lin_reg0.m

Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG DEPT.

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



Gradient descent

Eps=0.01

Real b=-0.5

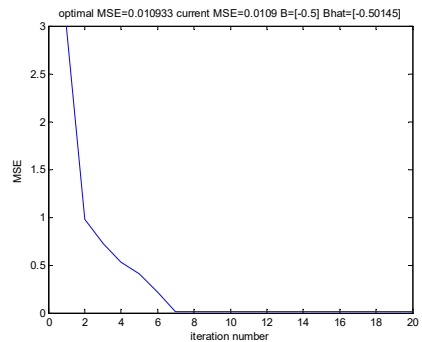
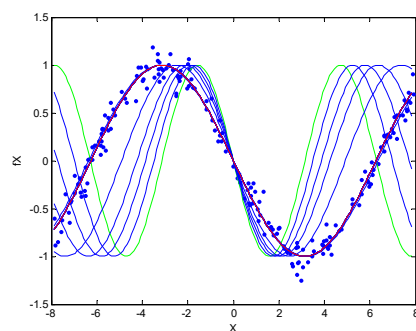
initial b=-1

Non_lin_reg0.m

Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG DEPT.

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



Newton raphson

Real b=-0.5

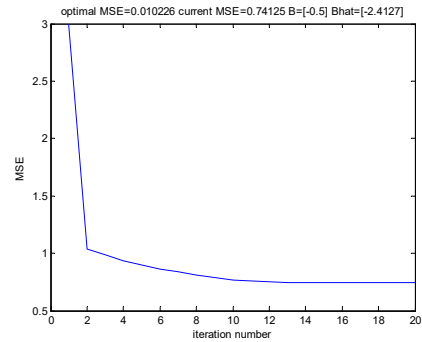
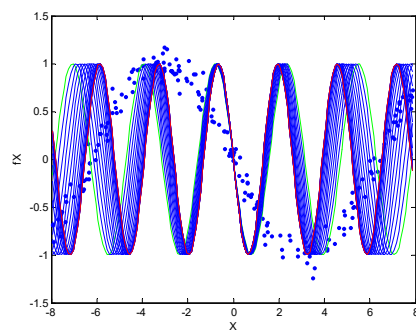
initial b=-1

Non_lin_reg0.m

Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG DEPT.

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



Newton raphson

Real $b = -0.5$

initial $b = -2$

Non_lin_reg0.m

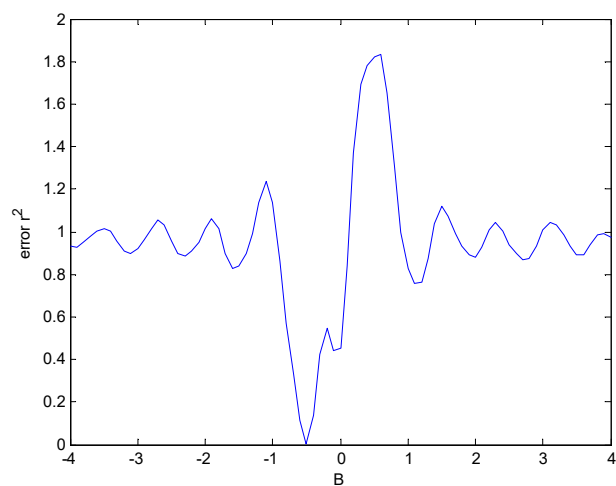
Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG. DEPT.

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

b_1 vs. error

- $B_1 = -0.5$



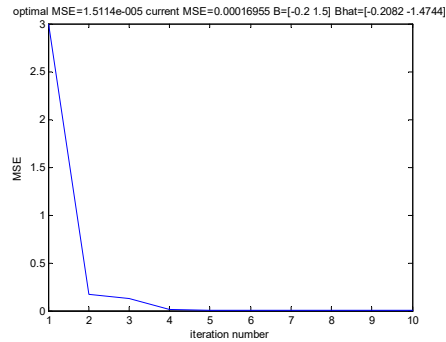
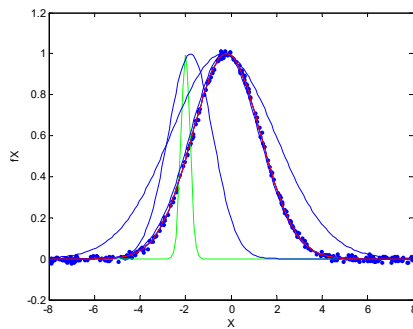
Non_lin_reg1.m

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

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



Gradient descent

Eps=0.1

Real B=[-0.2 ; 1.5]

initial B= [-2; -0.2]

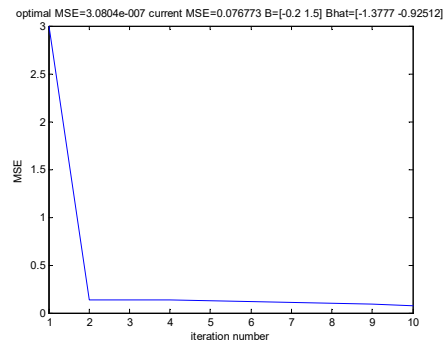
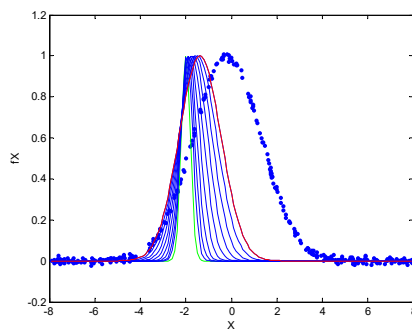
Non_lin_reg2.m

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

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



Gradient descent

Eps=0.01

Real B=[-0.2 ; 1.5]

initial B= [-2; -0.2]

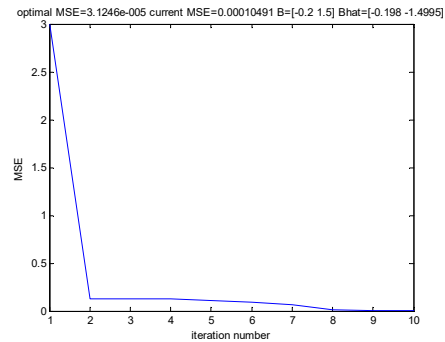
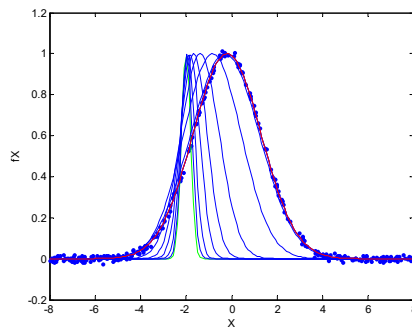
Non_lin_reg2.m

Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG. DEPT.

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

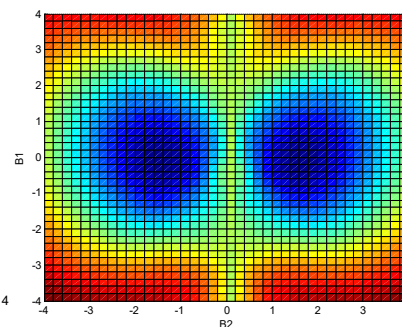
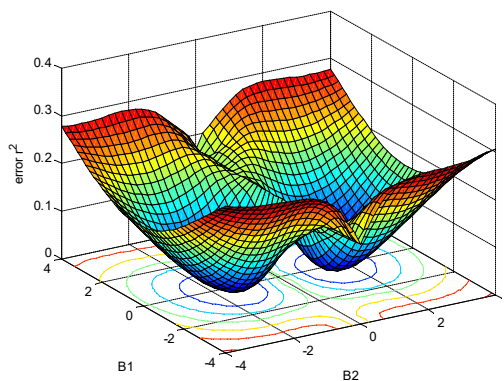
Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

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

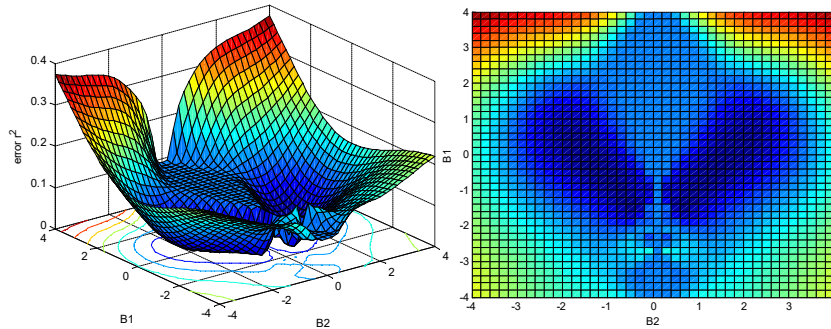
YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG DEPT.

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

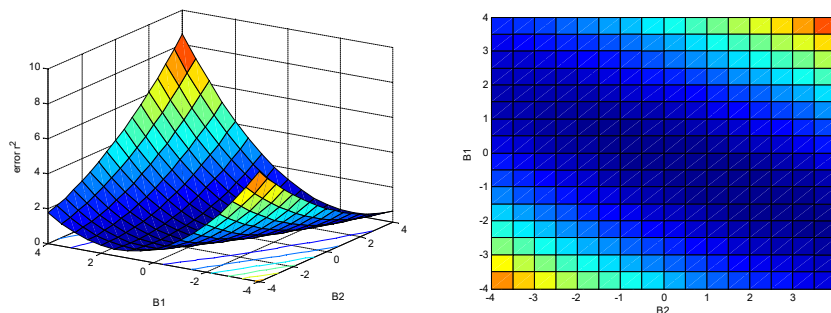
Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

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

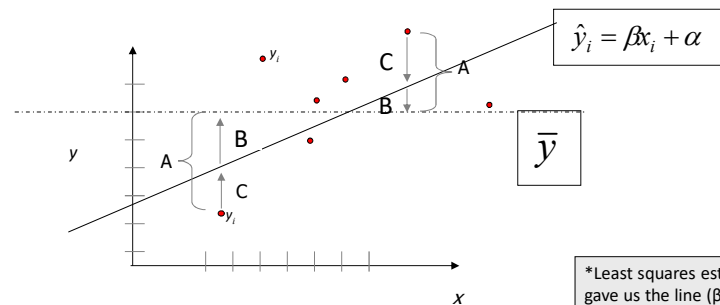
Mehmet Fatih AMASYALI Optimization Techniques Lecture Notes

YILDIZ TECHNICAL UNIVERSITY COMPUTER ENG DEPT.

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



*Least squares estimation gave us the line (β) that minimized C^2

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

A^2 B^2 C^2

SS_{total}
Total squared distance of
observations from naive mean of y
Total variation

SS_{reg}
Distance from regression line to naive mean of y
Variability due to x (regression)

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

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

References

- <http://www.columbia.edu/~so33/SusDev/Lecture3.pdf>
- http://www.msu.edu/~fuw/teaching/Fu_Ch11_linear_regression.ppt
- http://www.holehouse.org/mlclass/10_Advice_for_applying_machine_learning.html
- <http://www.imm.dtu.dk/~pcha/LSDF/NonlinDataFit.pdf>
- <http://fourier.eng.hmc.edu/e176/lectures/NM/node34.html>
- <http://fourier.eng.hmc.edu/e176/lectures/NM/node21.html>
- <http://www.kenbenoit.net/courses/ME104/logmodels2.pdf>
- <http://stattrek.com/regression/linear-transformation.aspx>
- [http://en.wikipedia.org/wiki/Data_transformation_\(statistics\)](http://en.wikipedia.org/wiki/Data_transformation_(statistics))