I3344
Numerical Simulation & Modelling

Ahmad Fadlallah

Lebanese University / Faculty of Science - I

Interpolation and Extrapolation

General Introduction

General Introduction

- In many cases, we know the values of a function $f(x)$ at a set of points $x_1 < x_2 < ... < x_N$, but we don't have the analytic expression of the function that lets us **calculate its value at an arbitrary point**.
- We will try to **estimate** $f(x)$ for arbitrary x by "drawing" a curve through the x_i and sometimes beyond them.
- Usually known as "**Curve Fitting**"

General Introduction (cont'd)

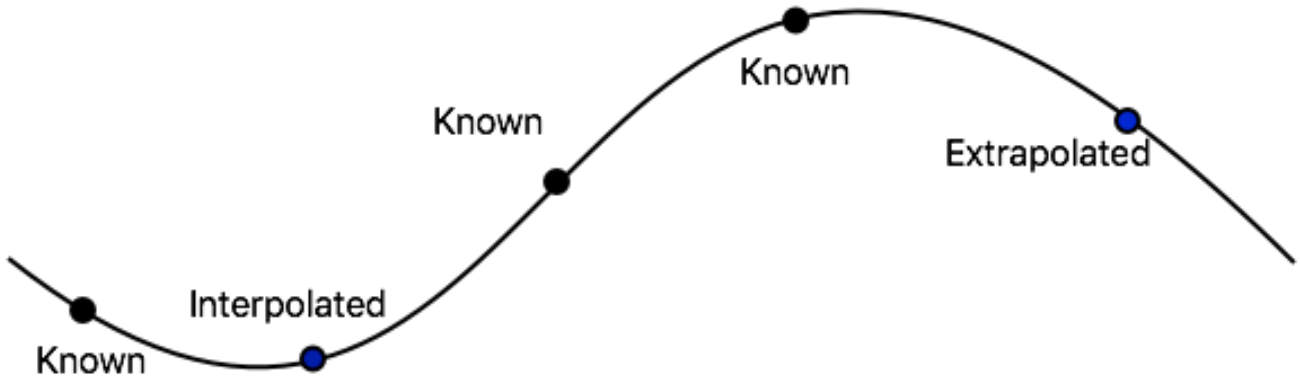
BUT why do we need such things?

- Simulation
- Computer Graphics
- Modeling/ Forecast
- Artificial Intelligence
- Machine Learning
- Several research and application domains
- etc.

General Introduction (cont'd)

The procedure of estimating the value of f(x) is called:

- **Interpolation** for $x \in [x_1, x_N]$
- **Extrapolation** for $x \notin [x_1, x_N]$



The form of the function that approximates the set of points **should be a convenient one** and should be **applicable to a general class of problems**. Different estimation methods are available

Linear Regression

Introduction

Polynomial approximation is usually used to find a polynomial function of order n (for n points) :

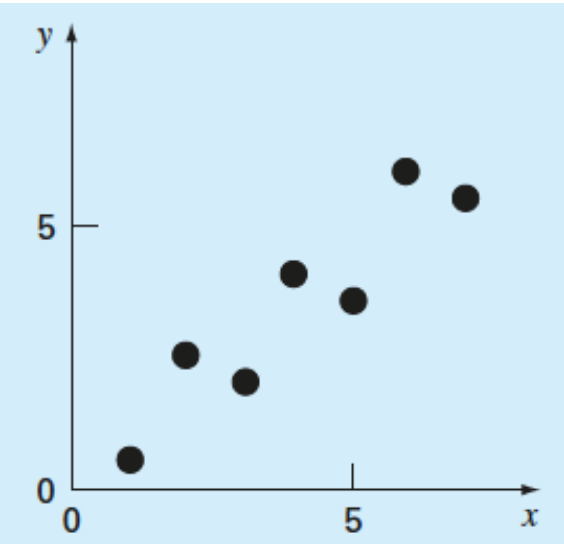
$$f(x) = a_{n-1}x^{n-1} + ... + a_0$$

- Finding coefficients a_i **requires** resolving a system of n equations in n unknowns
- Different methods were proposed for polynomial approximations
  - Lagrange's Polynomial
  - Hermite Polynomial
  - Taylor Polynomial
  - Cubic Splines
  - etc.
- The resulting curve will pass through all the n points

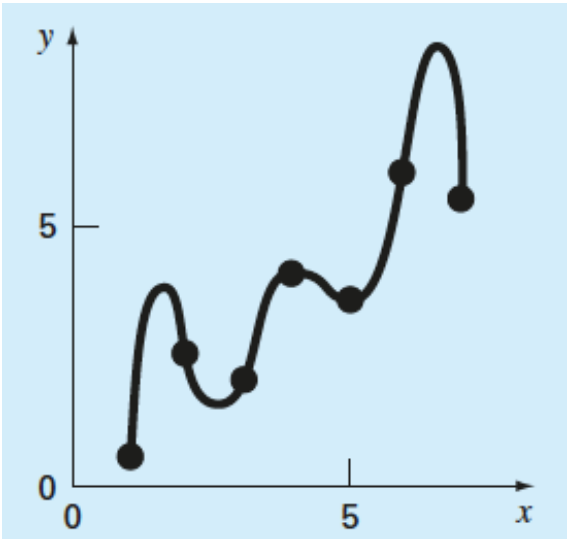
Introduction (cont'd)

- When data exhibit a significant degree of error or noise, the strategy is to **derive a single curve** that represents the general trend of these data.
- Because any individual data point may be incorrect, *no effort is made to intersect every point*.
- The curve is **designed to follow the pattern** of the points taken as a group.
- One approach of this nature is called **Least-Squares Regression**

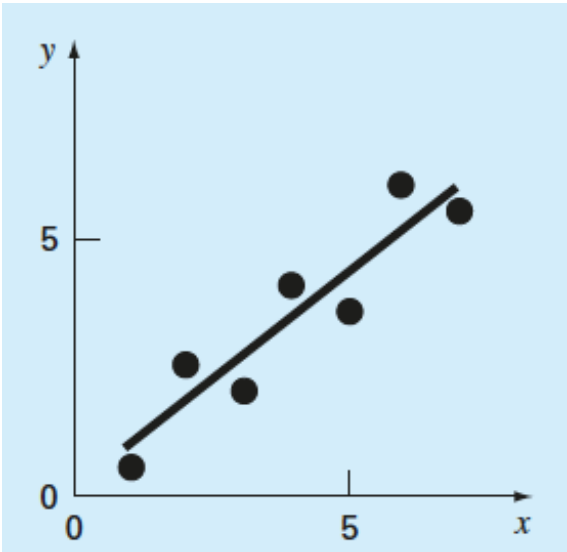
Introduction (cont'd)



Data



Polynomial Fit



Least-Square Fit

Linear Regression - Definition

- Simplest example of a **least-squares approximation** is fitting a straight line to a set of paired observations:

$$(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$$

- The mathematical expression for the straight line is:

$$y = a_0 + a_1 x + e$$

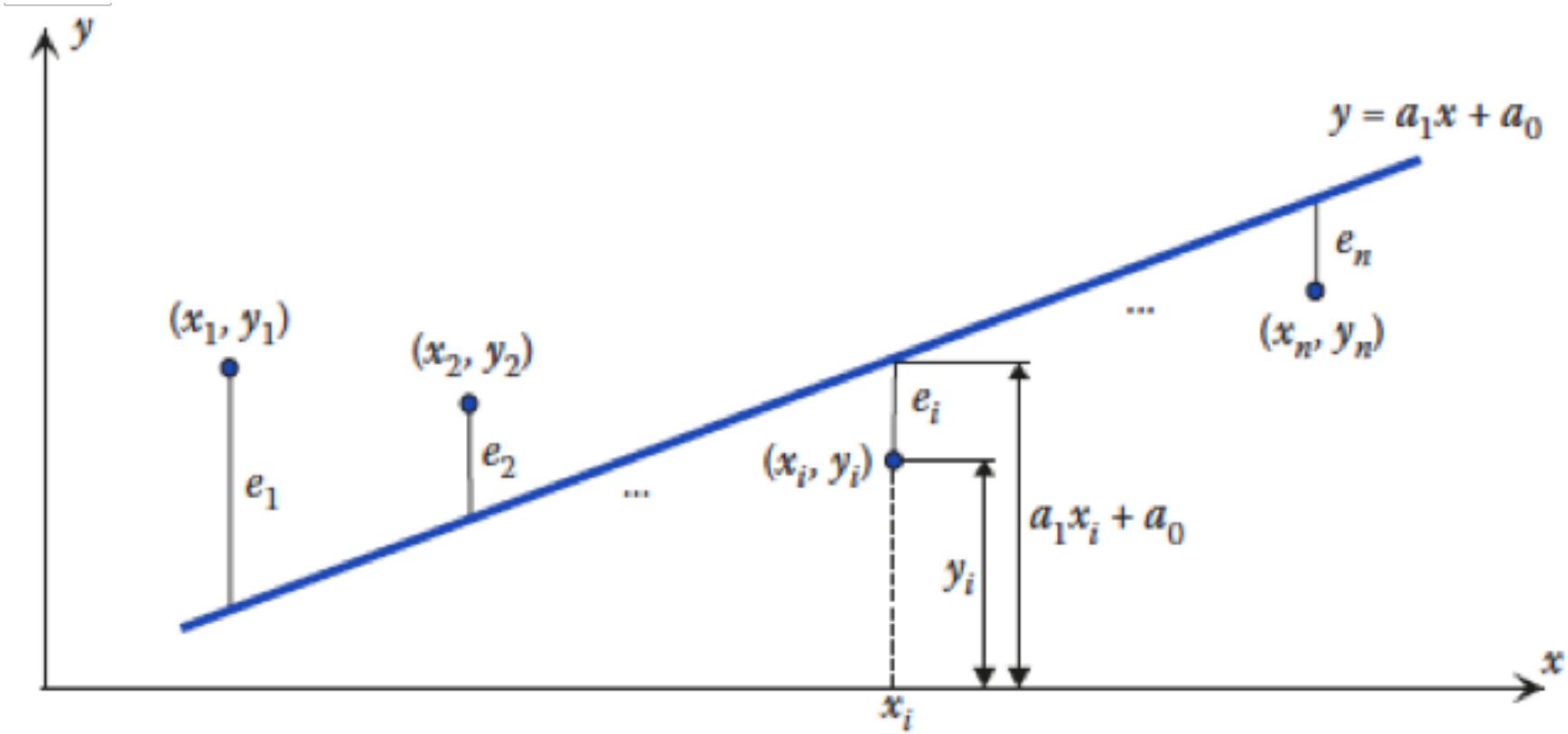
  - a_0 and a_1 are coefficients representing the **intercept** and the **slope** respectively
  - e is the **error**, or **residual**, between the model and the observations

Linear Regression - Definition (cont'd)

The error (e) can be represented by rearranging the equation as:

$$e = y - a_0 - a_1 x = y - (a_0 + a_1 x) = y - \hat{y}$$

The error (e) is the discrepancy between the true value of y and the approximate value $\hat{y}$ by the linear equation



Criteria for a Best Fit

First Approach:

Minimize the sum of residual errors

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)$$

where n is the total number of points

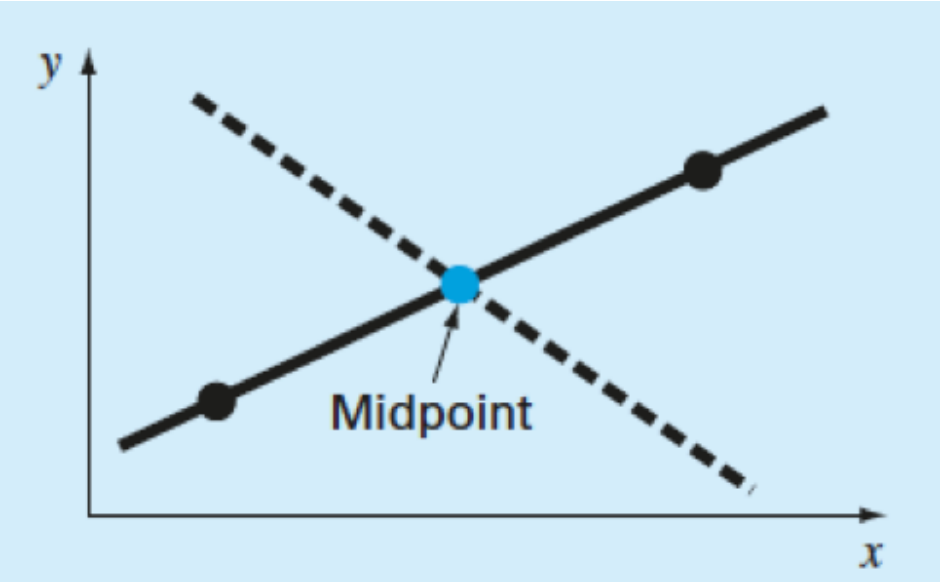
Criteria for a Best Fit

First Approach:

Minimize the sum of residual errors

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)$$

where n is the total number of points



NOT A Good Choice
since the errors might cancel

Criteria for a Best Fit (cont'd)

Second Approach:

Minimize the sum of absolute values of residual errors

$$\sum_{i=1}^{n} |e_i| = \sum_{i=1}^{n} |(y_i - a_0 - a_1 x_i)|$$

where n is the total number of points

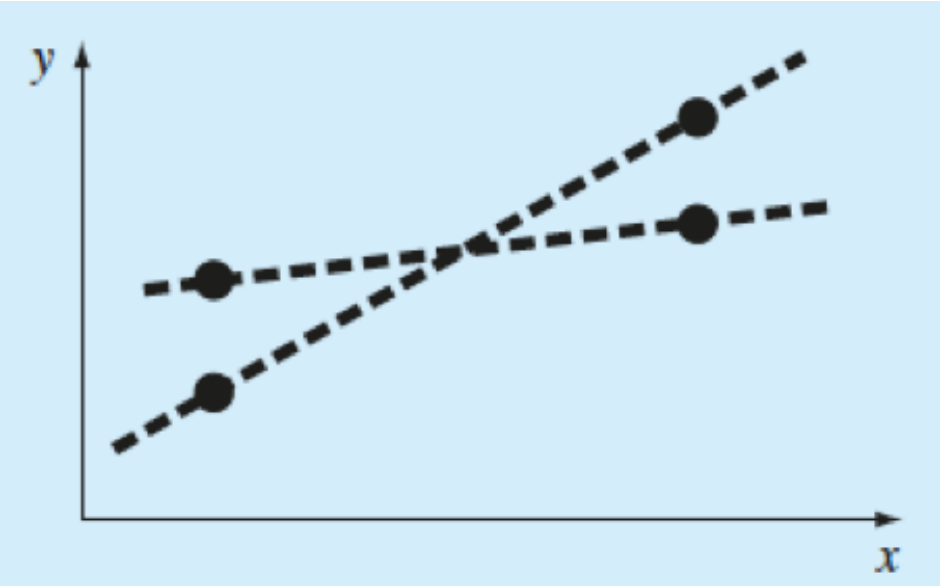
Criteria for a Best Fit (cont'd)

Second Approach:

Minimize the sum of absolute values of residual errors

$$\sum_{i=1}^{n} |e_i| = \sum_{i=1}^{n} |(y_i - a_0 - a_1 x_i)|$$

where n is the total number of points



NOT A Good Choice

Criteria for a Best Fit (cont'd)

Third Approach:
Minimize the Sum of Squares of residual Errors (SSE) or S_r

$$S_r = \sum_{i=1}^{n} |e_i^2| = \sum_{i=1}^{n} (y_i - \hat{y})^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

where n is the total number of points

This approach allows to overcome the shortcomings of the previous approaches

Least-Squares Fit of a Straight Line

$$S_r = \sum_{i=1}^{n} |e_i^2| = \sum_{i=1}^{n} (y_i - \hat{y})^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

Differentiate the equation with respect to each coefficient a_0 and a_1

$$\frac{\delta S_r}{\delta a_0} = -2 \sum_i (y_i - a_0 - a_1 x_i)$$

$$\frac{\delta S_r}{\delta a_1} = -2 \sum_i [(y_i - a_0 - a_1 x_i) x_i]$$

Minimize $S_r \Rightarrow$

$$\frac{\delta S_r}{\delta a_0} = \frac{\delta S_r}{\delta a_1} = 0$$

Least-Squares Fit of a Straight Line (cont'd)

$$0 = \sum (y_i - a_0 - a_1 x_i) = \sum y_i - \sum a_0 - \sum a_1 x_i$$

$$0 = \sum [(y_i - a_0 - a_1 x_i) x_i] = \sum y_i x_i - \sum a_0 x_i - \sum a_1 x_i^2$$

$$\sum a_0 = n a_0$$

$$n a_0 + (\sum x_i) a_1 = \sum y_i$$

$$(\sum x_i) a_0 + (\sum x_i^2) a_1 = \sum x_i y_i$$

$$a_1 = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{n(\sum x_i^2) - (\sum x_i)^2}, a_0 = \frac{(\sum x_i)^2(\sum y_i) - (\sum x_i)(\sum x_i y_i)}{n(\sum x_i^2) - (\sum x_i)^2}$$

Least-Squares Fit of a Straight Line (cont'd)

Summary

$$a_1 = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{n(\sum x_i^2) - (\sum x_i)^2}$$

$$a_0 = \overline{y} - a_1 \overline{x}$$

Exercise

Fit a straight line to the following x and y values

x_i	y_i
1	0.5
2	2.5
3	2.0
4	4.0
5	3.5
6	6.0
7	5.5

$$a_1 = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{n(\sum x_i^2) - (\sum x_i)^2}$$

$$a_0 = \overline{y} - a_1 \overline{x}$$

Exercise

Fit a straight line to the following x and y values

x_i	y_i
0.2	8.2
0.4	8.4
0.6	8.5
0.8	8.6
1.0	8.8
1.2	8.7

$$a_1 = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{n(\sum x_i^2) - (\sum x_i)^2}$$

$$a_0 = \overline{y} - a_1 \overline{x}$$

Background - Statistics

Consider a set of n values $y_1, ..., y_n$

- **Mean**: $\overline{y} = \frac{\sum y_i}{n}$
- **Standard Deviation**: allows to express the **dispersion** of the values **around their mean** i.e. the average distance between the values and their mean.

  $\sigma = \sqrt{\nu}$ where the **variance** $\nu = \frac{[\sum_{i=1}^{n} (x_i - \bar{x})^2]}{n}$
- **Correlation**: Describe the **relation between two variables**

  $r = cor(x, y) = \frac{cov(x, y)}{(\sigma_x * \sigma_y)}$

  $cov(x, y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n}$

Background - Statistics

The correlation coefficient $r \in [-1, 1]$ in contrast to the covariance that is influenced by the range of values of x and y.

Three types of correlation:

- **Positive correlation**: x $\nearrow \Rightarrow$ y $\nearrow$
- **Negative correlation**: x $\nearrow \Rightarrow$ y $\searrow$
- **Zero correlation**: No relationship exists between x and y.

The closer the value is to 1 or -1, the stronger the link.

Linear Regression - Revisited

$$a_1 = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{n(\sum x_i^2) - (\sum x_i)^2}, a_0 = \frac{(\sum x_i)^2(\sum y_i) - (\sum x_i)(\sum x_i y_i)}{n(\sum x_i^2) - (\sum x_i)^2}$$

$$a_1 = r * \frac{\sigma_y}{\sigma_x}, \ a_0 = \bar{y} - a_1 \bar{x}$$

Linear Regression - Revisited

$$a_1 = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{n(\sum x_i^2) - (\sum x_i)^2}, a_0 = \frac{(\sum x_i)^2(\sum y_i) - (\sum x_i)(\sum x_i y_i)}{n(\sum x_i^2) - (\sum x_i)^2}$$

$$a_1 = r * \frac{\sigma_y}{\sigma_x}, \ a_0 = \bar{y} - a_1 \bar{x}$$

Exercise: **Prove** $a_1 = r * \frac{\sigma_y}{\sigma_x}$

$$\sigma = \sqrt{\nu}, \ \nu = \frac{[\sum_{i=1}^{n} (x_i - \bar{x})^2]}{n}$$

$$r = cor(x, y) = \frac{cov(x, y)}{(\sigma_x * \sigma_y)}$$

$$cov(x, y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n}$$

Quantification of Error of Linear Regression

- Any line other than the one computed results in a larger Sum of the Squares of the Residuals (SSE)
- The line is **unique** and in terms of our chosen criterion is a "best" line through the points.
- How to calculate error in Linear Regression?

Suggestion: Using the Sum of the Squares of the Residuals (SSE)

Exercise

Using previously calculated coefficients, calculate the SSE

x_i	y_i
1	0.5
2	2.5
3	2.0
4	4.0
5	3.5
6	6.0
7	5.5

$$S_r = SSE = \sum_{i=1}^{n} |e_i^2| = \sum_{i=1}^{n} (y_i - \hat{y})^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

Quantification of Error of Linear Regression

How to evaluate the accuracy of a_1 and a_0?

- The value of SSE does not make it possible to evaluate the accuracy of these two coefficients.
- Example: an error of 2.5 on a set of data belonging to the interval [1000, 1500] is negligible in comparison with the same value on another set belonging to the interval [0,1].

Quantification of Error of Linear Regression

How to evaluate the accuracy of a_1 and a_0?

- The value of SSE does not make it possible to evaluate the accuracy of these two coefficients.
- Example: an error of 2.5 on a set of data belonging to the interval [1000, 1500] is negligible in comparison with the same value on another set belonging to the interval [0,1].

Solution: **Normalize the Error**

Quantification of Error of Linear Regression (cont'd)

Method 1: Use Normalized Root-Mean-Square Error (NRMSE)

$$MSE = \frac{SSE}{n}, \quad RMSE = \sqrt{\frac{SSE}{n}}, \quad NRMSE = \frac{RMSE}{\bar{y}}$$

- NRMSE close to 0 $\Rightarrow$ **Perfect model**
- NRMSE close to 1 $\Rightarrow$ **Null model**

NRMSE might be expressed as follows:

$$NRMSE = \frac{RMSE}{(Max - Min)}$$

Exercise

Calculate the NRMSE and Evaluate

x_i	y_i
1	0.5
2	2.5
3	2.0
4	4.0
5	3.5
6	6.0
7	5.5

x_i	y_i
0.2	8.2
0.4	8.4
0.6	8.5
0.8	8.6
1.0	8.8
1.2	8.7

$$MSE = \frac{SSE}{n}, \quad RMSE = \sqrt{\frac{SSE}{n}}, \quad NRMSE = \frac{RMSE}{\bar{y}}$$

$$S_r = SSE = \sum_{i=1}^{n} |e_i^2| = \sum_{i=1}^{n} (y_i - \hat{y})^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

Quantification of Error of Linear Regression (cont'd)

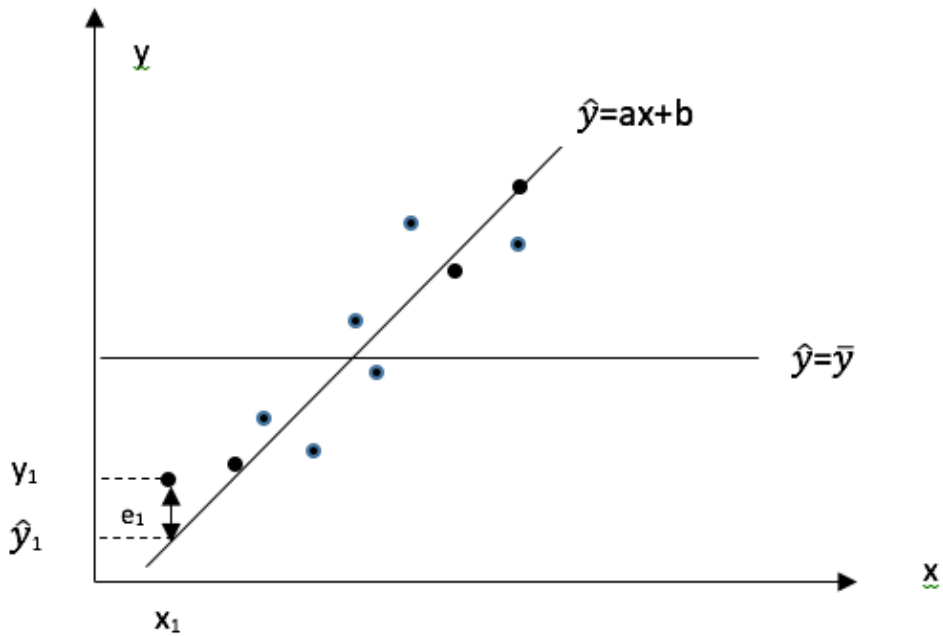
Method 2: Compare the approximate model $\hat{y} = ax + b$ with the Null model $\hat{y} = \bar{y}$ (where x does not intervene)

We introduce the measure $R^2 = 1 - SSE/TSS$

SSE = **Sum of Square Errors**
TSS = **Total Sum of Squares**

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

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



close to 1 $\Rightarrow$ Perfect Model
close to 0 $\Rightarrow$ Null Model

SSE $<$ TSS since in worst case scenario the model found is the null model

Exercise

Calculate R^2 and Evaluate

x_i	y_i
1	0.5
2	2.5
3	2.0
4	4.0
5	3.5
6	6.0
7	5.5

x_i	y_i
0.2	8.2
0.4	8.4
0.6	8.5
0.8	8.6
1.0	8.8
1.2	8.7

$$R^2 = 1 - SSE/TSS, \quad SSE = \sum_{i=1}^{n} (y_i - \hat{y})^2, \quad TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

K-Nearest Neighbors

KNN Algorithm

K-Nearest Neighbors (KNN) can be used in case of non-linear relations

Given x_{n+1} of a new point P_{n+1}, find the approximate value $\hat{y}_{n+1}$

Algorithm

1. Choose an integer $k \in [1, n]$ where n is the number of points.
2. Calculate the **Euclidean distance** $\sqrt{(x_{n+1} - x_i)^2}$ between the new point and the n points according to the abscissa
3. Choose the **k nearest points**: $P_{i_1}, P_{i_2}, ..., P_{i_k}$
4. $\hat{y}_{n+1} = \dfrac{(y_{i_1} + y_{i_2} + ... + y_{i_k})}{k}$

How to choose k?

A good rule is $k = \sqrt{n}$

Exercise

Apply KNN to the approximation for the value x=4.4 and for the values k=3, 4 and 5

x_i	y_i
1	0.5
2	2.5
3	2.0
4	4.0
5	3.5
6	6.0
7	5.5

K-fold Cross Validation

Introduction

- **K-fold Cross Validation** is a model validation technique for assessing *how the results of a statistical analysis will* **generalize** *to an independent data set*.
- It is mainly used in settings where **the goal is prediction**, and one wants to **estimate** how **accurately** a **predictive model will perform in practice**.
- In a **prediction problem**, a model is usually given a dataset of known data on which training is run (**training dataset**), and a dataset of unknown data against which the model is tested (**validation dataset or testing set**)

Cross Validation

- The goal of **cross validation** is to **define** a validation set, in order to:
  - Limit problems like overfitting
  - Give an insight on how the model will generalize to an independent dataset in a real case scenario for instance
- Various cross validation techniques
  - Test-set validation or holdout method
  - **K-fold cross-validation**
  - Leave-One-Out Cross-Validation (LOOCV)

Test-Set Validation / Holdout Method

- Simplest kind of cross validation.
- The **data set** is separated into two sets: the **training set** and the **testing set**.
- The **function "approximator"** fits a function using the training set only.
- Then the function "approximator" is asked to predict the output values for the data in the testing set (it has never seen these output values before).

How K-Fold Cross-Validation Works?

In k-fold cross-validation, the original sample is randomly partitioned into **k equal sized subsamples**.

Of the k subsamples, **a single subsample is retained as the validation data** for testing the model, and **the remaining k1 subsamples** are used as **training data**.

The **cross-validation process is then repeated k times** (the folds), with each of the k subsamples used exactly once as the validation data.

The **k results** from the folds can then be **averaged** to produce a single estimation.

How K-Fold Cross-Validation Works?

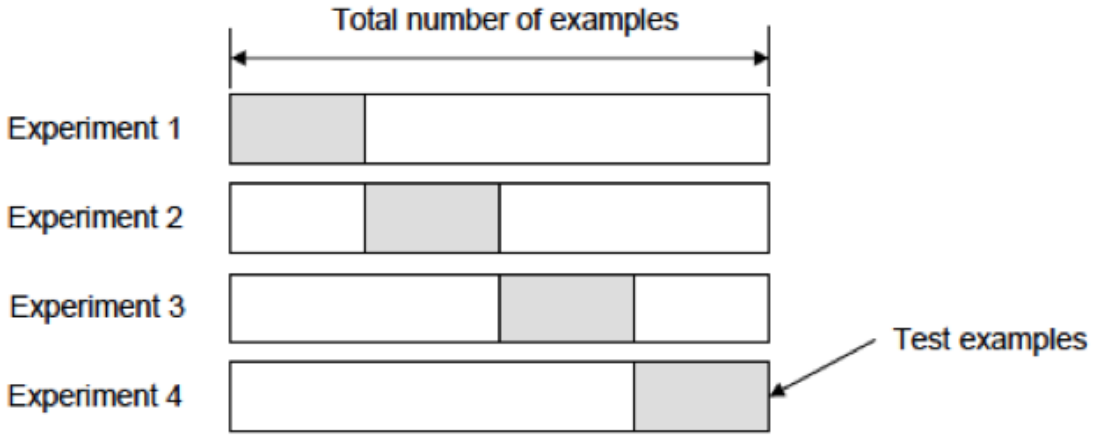
Algorithm

Randomly divide the set E into k separate subsets (usually k = 10): $E_1, ..., E_$

For i:1$\rightarrow$ k

Learning set: $E_j, 1 \leq j \leq k, j \neq i$

Validation set: E_i



$$CV = \frac{\sum_{i=1}^{k} NRMSE_i}{k}$$

$$NRMSE_i = \frac{\sqrt{\frac{\sum_j (y_j - \bar{y}_j)^2}{n/k}}}{\bar{y}}$$

Cross Validation of Models

Cross-validation makes it possible to compare two different models.

One model is considered **more effective** than another if its **CV is smaller**.

Leave-One-Out Cross-Validation (LOOCV)

- K-fold cross validation taken to its logical extreme
- K = N (the number of data points in the set)
- N separate times, the function approximator is trained on all the data except for one point and a prediction is made for that point.

Exercise

Compare KNN (K=3) and Linear Regression using K-fold cross validation (K=3)

x_i	y_i
1	0.5
2	2.5
3	2.0
4	4.0
5	3.5
6	6.0
7	5.5

$$MSE = \frac{SSE}{n},\ RMSE = \sqrt{\frac{SSE}{n}},\ NRMSE = \frac{RMSE}{\bar{y}}$$

$$CV = \frac{\sum_{i=1}^{k} NRMSE_i}{k}$$

$$a_1 = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{n(\sum x_i^2) - (\sum x_i)^2}$$

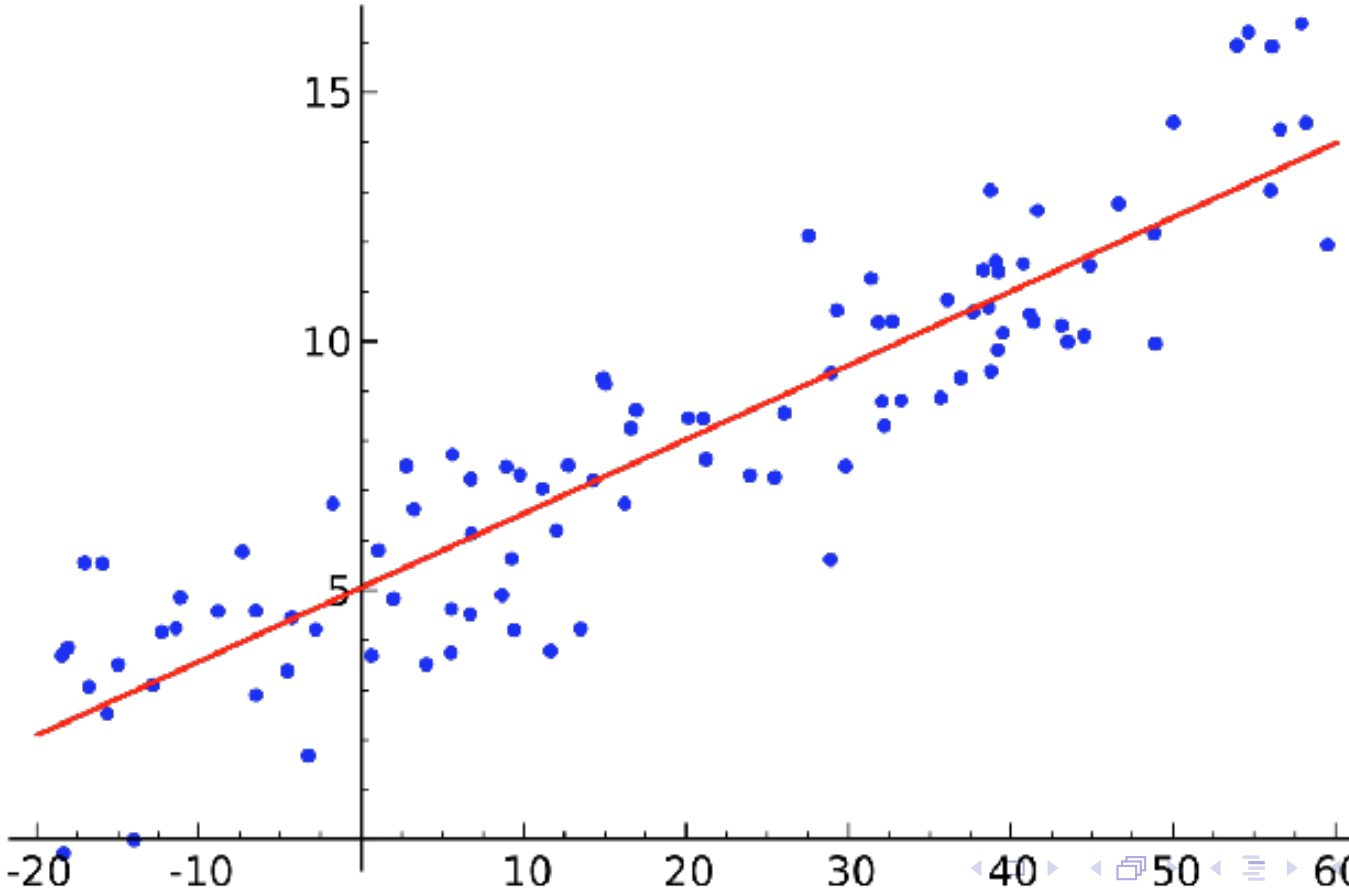
$$a_0 = \frac{(\sum x_i^2))(\sum y_i) - (\sum x_i)(\sum x_i y_i)}{n(\sum x_i^2) - (\sum x_i)^2}$$

Multiple Linear Regression

Introduction

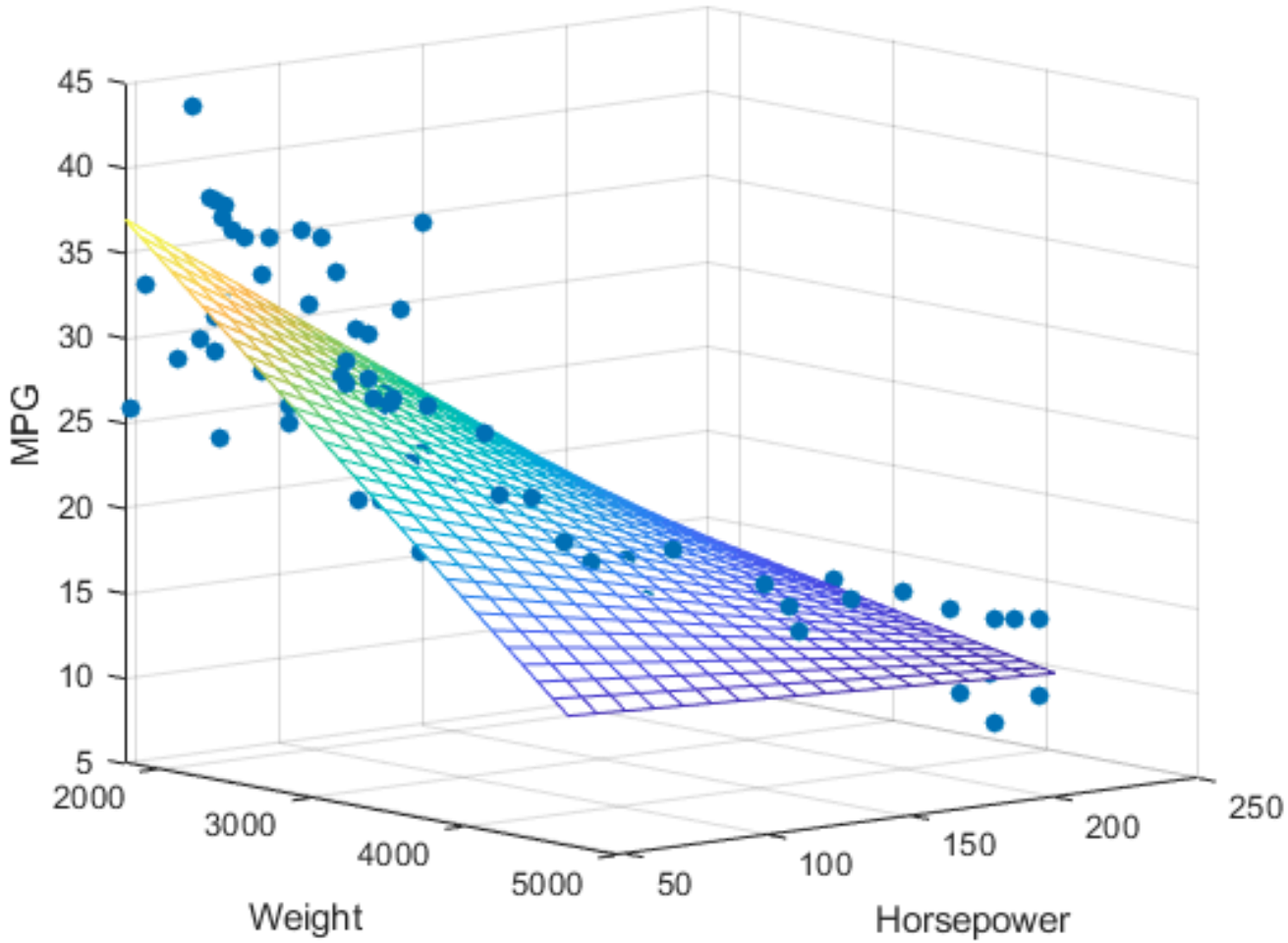
Simple Linear Regression: Prediction of variable y depends on one predictor (variable)

$\hat{y} = f(x) = ax + b$



Multiple Linear Regression - Definition

Prediction of variable y depends on two or more **predictors** x_1, x_2, ... , x_p



Multiple Linear Regression - Definition (cont'd)

Prediction of variable y depends on two or more **predictors** x_1, x_2, ... , x_p

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ... + \beta_p x_p + \epsilon$$

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ... + \beta_p x_p$$

$$\hat{y}_i = \sum_{j=1}^{p} \beta_j x_{i,j} + \beta_0,$$

$2 \leq$ p $\leq$ m-1 / m is the # of variables

Meaning of coefficients β_i

- single unit change in $x_i \Rightarrow \beta_i$ change in y ($\approx$ slope concept)
- positive β_i : $x_i \uparrow \Rightarrow$ y $\uparrow$
- negative β_i : $x_i \uparrow \Rightarrow$ y $\downarrow$

Multiple Linear Regression - Objectives

- Analyze how y is affected by the different variables/predictor/factors
- Determine the relation between one or two variables (x_i) (**key predictors**), the rest of predictors are added to have a more accurate model

Multiple Linear Regression - Example

Housing prices

- Dependent variable: Home sale price
- **Key predictor variable**: square footage
- Other factors/predictors:
  - Median neighborhood income
  - Age of the home
  - size of the lot
  - etc.

Provide more accurate coefficients for the **key predictor** variable

- Suppose you didn't consider the factor of house age

Choice of Predictors

- Adding more variables will change the coefficients of other predictor variables including the key predictor
- Picking the "right" variables in MLR is very important
- Bad choice of predictor variables lead to a wrong prediction $\Rightarrow$ wrong conclusions
  - Dummy variables might appear (erroneously) significant
- Avoid picking variables that are "overlapping"

Calculating Coefficients

How to get a "good" estimate of y?

Least Square method $\Rightarrow$ Be Wrong as little as possible (Similar to Simple Linear Regression)

$$\begin{bmatrix} y_1 \\ \dots \\ y_n \end{bmatrix} = \begin{pmatrix} 1 & x_{1,1} & \dots & x_{1,p} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n,1} & \dots & x_{n,p} \end{pmatrix} x \begin{bmatrix} \beta_0 \\ \dots \\ \beta_p \end{bmatrix} + \begin{bmatrix} e_1 \\ \dots \\ e_n \end{bmatrix}$$

$$Y = X * \beta + e$$

Choose β_i to minimize SSE

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{i,j})^2$$

$$\text{Minimizing SSE} \Rightarrow \frac{d(SSE)}{d\beta} = 0 \Rightarrow \boxed{\beta = (X^t X)^{-1} X^t Y}$$

Calculating Coefficients - Example

Find $\hat{y} = f(x_1, x_2)$

y	x_1	x_2
5	1	2
8	2	3
9	3	2
12	4	3

$$\mathsf{Y} = \mathsf{X}\,\beta + \mathsf{e}$$

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

Calculating Coefficients - Example

Find $\hat{y} = f(x_1, x_2)$

y	x_1	x_2
5	1	2
8	2	3
9	3	2
12	4	3

$$\mathrm{Y} = \mathrm{X}\,\beta + \mathrm{e}$$
$$\beta = (X^t X)^{-1} X^t Y$$

$$\mathrm{Y} = \begin{bmatrix} 5 & 8 & 9 & 12 \end{bmatrix}, \mathrm{X} = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 1 & 3 & 2 \\ 1 & 4 & 3 \end{bmatrix}$$

Calculating Coefficients - Example

Find $\hat{y} = f(x_1, x_2)$

y	x_1	x_2
5	1	2
8	2	3
9	3	2
12	4	3

$$\mathsf{Y} = \mathsf{X}\,\beta + \mathsf{e}$$
$$\beta = (X^t X)^{-1} X^t Y$$

$$(X^t X)^{-1} X^t = \begin{bmatrix} 1.5 & -1 & 1.5 & -1 \\ -0.25 & -0.25 & 0.25 & 0.25 \\ -0.25 & 0.75 & -0.75 & 0.25 \end{bmatrix}$$

Calculating Coefficients - Example

Find $\hat{y} = f(x_1, x_2)$

y	x_1	x_2
5	1	2
8	2	3
9	3	2
12	4	3

$$\mathsf{Y} = \mathsf{X}\,\beta + \mathsf{e}$$
$$\beta = (X^t X)^{-1} X^t Y$$

$$(X^t X)^{-1} X^t = \begin{bmatrix} 1.5 & -1 & 1.5 & -1 \\ -0.25 & -0.25 & 0.25 & 0.25 \\ -0.25 & 0.75 & -0.75 & 0.25 \end{bmatrix}$$

$$\beta = (\beta_0, \beta_1, \beta_2) = (1, 2, 1)$$

$$\hat{y} = (5, 8, 9, 12),\ SSE = 0$$

Choosing the predictors

First Method: Combination

for m $= 3$ (y, x_1, x_2, there are four possibilities

- $y = f(x_1, x_2)$
- $y = f(x_1)$
- $y = f(x_2)$
- $y = \beta_0$

m variables $\Rightarrow 2^{m-1}$ possibilities $\Rightarrow$ too high complexity when n is $\gg$

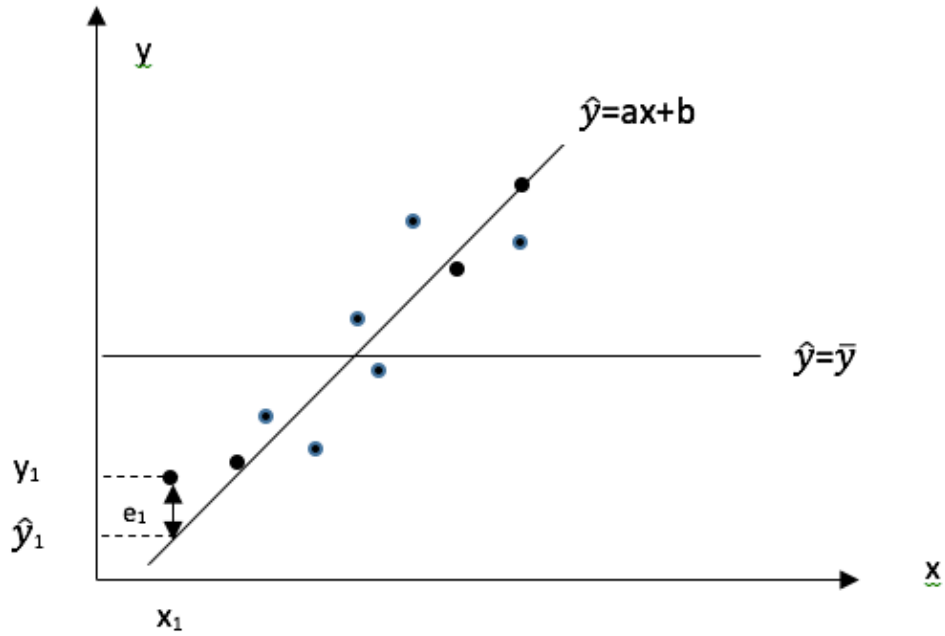
Model Evaluation - Background

$R^2 = (1 - SSE/TSS) \in [0, 1]$
SSE = **Sum of Square Errors**
TSS = **Total Sum of Squares**

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

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

higher $R^2 \Rightarrow$ better predictive model
close to 1 $\Rightarrow$ Perfect Model
close to 0 $\Rightarrow$ Null Model



SSE $<$ TSS since in worst case scenario the model found is the null model

Model Evaluation (cont'd)

Problem with R^2 : It always increases as you add new variables but this doesn't mean that our model is better!
Solution: Adjusted R^2

$$R_a^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1} = 1 - \frac{SSE}{TSS}\frac{n-1}{n-p-1}$$

Choosing the Predictors (cont'd)

Second Method: Direct Selection

- **Initialization:** Start with null model: $\hat{y} = \beta_0 = \bar{y}$
- **Step 1:** add one variable at a time and check the model which have the greatest Adj. R^2
- **Step 2:**
  - if Adj. R^2 of the new model=1 $\Rightarrow$ perfect model
  - Else if Adj. R^2 is less or equal to that of the previous model $\Rightarrow$ take the previous model
  - Else go to step 1

Exercise

Choose the "best" predictor using the Direct selection Method

A	B	C	D
1	2	14	3
3	5	23	4
7	9	17	8
4	5	19	6

Exercise

Choose the "best" predictor using the Direct selection Method

Initialization : start with null model (adj. $R^2 = 0$)

A	B	C	D
1	2	14	3
3	5	23	4
7	9	17	8
4	5	19	6

Exercise

Choose the "best" predictor using the Direct selection Method

Initialization : start with null model (adj. $R^2 = 0$)

- **Iteration 1** :
  - A=f(B), Adj-R^2=0.973
  - A=f(C), Adj-R^2=-0.4
  - A=f(D), Adj-R^2=0.95
  - A=f(B) is the best model, since 0.973>0 continue

A	B	C	D
1	2	14	3
3	5	23	4
7	9	17	8
4	5	19	6

Exercise

Choose the "best" predictor using the Direct selection Method

Initialization : start with null model (adj. $R^2 = 0$)

- **Iteration 1** :
  - A=f(B), Adj-R^2=0.973
  - A=f(C), Adj-R^2=-0.4
  - A=f(D), Adj-R^2=0.95
  - A=f(B) is the best model, since 0.973>0 continue
- **Iteration 2**:
  - A=f(B,C) : Adj-R^2=0.94
  - A=f(B,D) : Adj-R^2=1

A	B	C	D
1	2	14	3
3	5	23	4
7	9	17	8
4	5	19	6

Exercise

Choose the "best" predictor using the Direct selection Method

Initialization : start with null model (adj. $R^2 = 0$)

- **Iteration 1** :
  - A=f(B), Adj-R^2=0.973
  - A=f(C), Adj-R^2=-0.4
  - A=f(D), Adj-R^2=0.95
  - A=f(B) is the best model, since 0.973>0 continue
- **Iteration 2**:
  - A=f(B,C) : Adj-R^2=0.94
  - A=f(B,D) : Adj-R^2=1

A=f(B,D) is the selected model (perfect model)

A	B	C	D
1	2	14	3
3	5	23	4
7	9	17	8
4	5	19	6