Introduction to Machine Learning

Prof.N.L.Bhanu Murthy

What is Learning?

"Gain knowledge or understanding of or skill in by study, instruction or experience"

- Webster

What is Learning?

"**Learning** is any process by which a system improves **performance** from **experience**."

- Herbert Simon

Researcher in

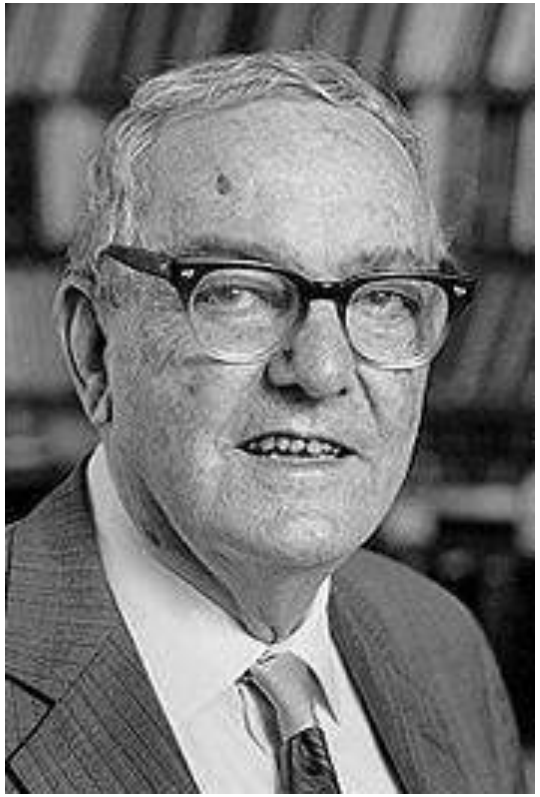
- ✓ Artificial Intelligence
- ✓ Cognitive psychology
- ✓ Computer science
- ✓ Economics
- ✓ Political science

Professor @

- ✓ Carnegie Mellon University
- ✓ University of California, Berkeley
- ✓ Illinois Institute of Technology

Awards:

- ✓ Turing Award, 1975
- ✓ Nobel Prize in Economics1978
- ✓ National Medal of Science1986
- ✓ von Neumann Theory Prize1988



1916 - 2001

What is Machine Learning?

Machine Learning is study of algorithms that

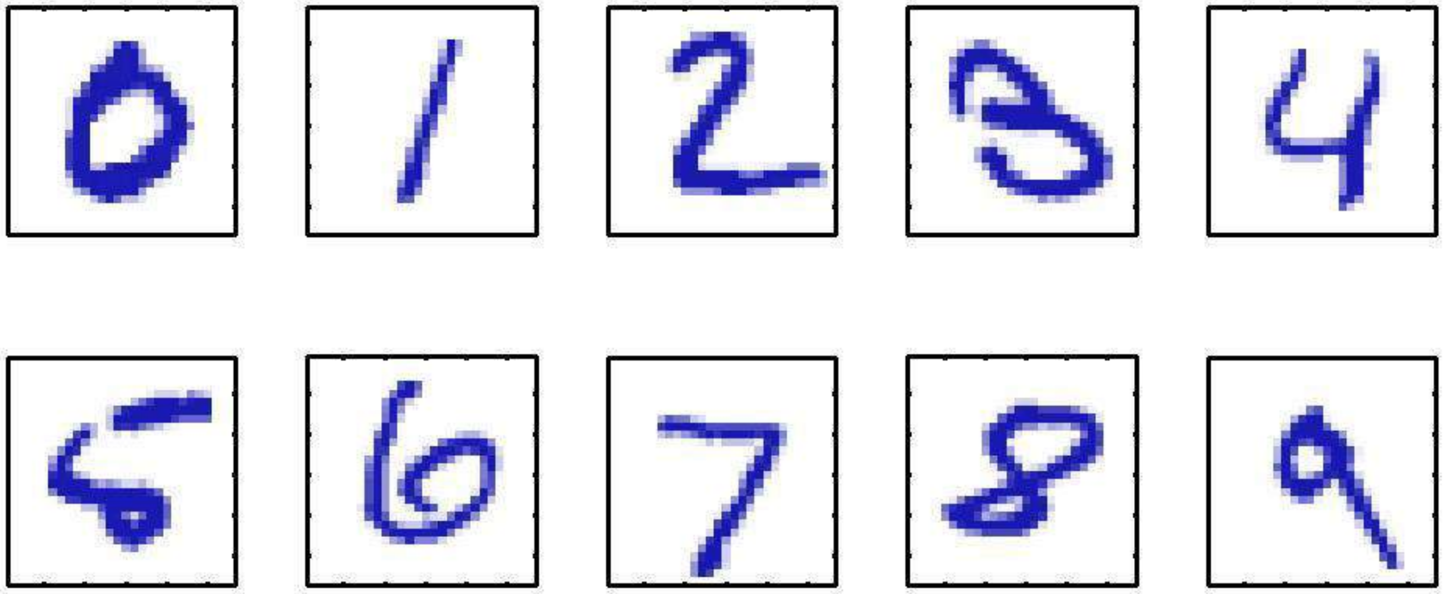
- improve their performance P
- at some task T
- with experience E

Tom Mitchell (1990)

Well-defined learning task: <P,T,E>

Example – Machine Learning

Handwritten Digit Recognition



T: Recognizing hand-written digits

E: Database of human-labeled images of handwritten digits

P: Percentage of digits correctly classified

Example – Machine Learning

Self-driving Vehicles

Little Ben, 60 miles of autonomous, safe, efficient driving

T: Driving on four-lane highways using vision sensors

P: Average distance traveled before a human-judged error

E: A sequence of images and steering commands recorded while observing a human driver.

Example – Machine Learning

- Learning to recognize spoken words (Lee, 1989; Waibel, 1989).
- Learning to classify new astronomical structures (Fayyad et al., 1995).
- Learning to play world-class backgammon (Tesauro 1992, 1995).
- Categorize email messages as spam or legitimate.

Machine Learning, a Magic?

No, more like gardening

- Seeds = Algorithms
- Nutrients = Data
- Gardener = You
- Plants = Programs

They said it!!

- ✓ “A breakthrough in machine learning would be worth ten Microsofts”

  - Bill Gates, Chairman, Microsoft

- ✓ Machine learning is the hot new thing” - John Hennessy, President, Stanford

- ✓ “Web rankings today are mostly a matter of machine learning”

  - Prabhakar Raghavan, Dir. Research, Yahoo

- ✓ “Machine learning is going to result in a real revolution” - Greg Papadopoulos, CTO, Sun

- ✓ “Machine learning is today’s discontinuity” - Jerry Yang, CEO, Yahoo

- ✓ “Machine learning is the next Internet” - Tony Tether, Director, DARPA

Future Prospects..

- Survey of AI researchers
  - AI will outperform humans in:
    - Translating languages 2024
    - Writing high-school essays 2026
    - Driving a truck 2027
    - Working in retail 2031
    - Writing a best-selling book 2049
    - Working as a surgeon 2053
    - Outperform humans in all tasks: 50% chance in 45 years
    - Automating all human jobs 120 years
  - Survey population: 2015 NIPS/ICML authors
    - Questions on AI capabilities (e.g. folding laundry, language translation), superiority at specific occupations (e.g. truck driver, surgeon), superiority over humans at all tasks.

12 IT skills that employers can't say no to

COMPUTER WORLD

1) **Machine learning**
2) Mobilizing applications
3) Wireless networking
4) Human-computer interface
5) Project management
6) General networking skills
7) Network convergence technicians
8) Open-source programming
9) Business intelligence systems
10) Embedded security
11) Digital home technology integration
12) .Net, C #, C ++, Java -- with an edge

Thank You!

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Introduction to Supervised Learning

Prof.N.L.Bhanu Murthy

Machine Learning – Examples (Employability Prediction)

Features / Attributes / Predictors

- ✓ CGPA
- ✓ Communication Skills
- ✓ Aptitude
- ✓ Programming Skills

S.No.	CGPA	Communication Skills	Aptitude	Programming Skills	Job Offered?
1	9.1	Average	Good	Excellent	Yes

Machine Learning – Examples (Employability Prediction)

Features / Attributes / Predictors

- ✓ CGPA
- ✓ Communication Skills
- ✓ Aptitude
- ✓ Programming Skills

S.No.	CGPA	Communication Skills	Aptitude	Programming Skills	Job Offered?
1	9.1	Average	Good	Excellent	Yes
2	8.4	Good	Good	Good	Yes
3	8.3	Poor	Average	Average	No
4	7.1	Average	Good	Average	No
5	8.2	Good	Excellent	Excellent	No

Machine Learning – Examples (Predicting price of a used car)

Features / Attributes / Predictors

- ✓ Brand
- ✓ Year (Mfg)
- ✓ Engine Capacity
- ✓ Mileage
- ✓ Distance travelled
- ✓ Cab?

S.No.	Brand	Year (Mfg)	Engine Capacity	Mileage	Distance travelled	Cab?	Price (in Rs.)
1.	Honda City ZX	2008	1100	10.5	45000	N	3,50,000
2							
3							
4							
5							

Machine Learning – Examples (Market Segmentation Study)

Features / Attributes / Predictors

- ✓ Family income
- ✓ # of visits in a month
- ✓ Average money spent in a month
- ✓ Zip code

Customers for a retailer may fall into

- ✓ two groups say big spenders and low spenders
- ✓ three groups say big spenders, medium spenders and low spenders
- ✓ Four groups, ….

S.No.	Zip Code	Family Income	# of visits in a month	Average Money Spent in a month
1	500078	11,50,000	4	8,000

Supervised Learning

Feature tuple: (CGPA, Communication Skills, Aptitude, Programming Skills)
Response / Target: Job Offered

Supervised Learning: Fit a model that relates response to the feature tuples, with the aim of accurately predicting the response for future observation or better understanding the relationship between response and features.

S.No.	CGPA	Communication Skills	Aptitude	Programming Skills	Job Offered?
1	9.1	Average	Good	Excellent	Yes
2	8.4	Good	Good	Good	Yes
3	8.3	Poor	Average	Average	No
4	7.1	Average	Good	Average	No
5	8.2	Good	Excellent	Excellent	No

Unsupervised Learning

Feature tuple: (Zip Code, Family Income, # of visits in a month, Average Money spent in a month)

Response / Target: None

Unsupervised Learning: To discover groups of similar examples within the data set

S.No.	Zip Code	Family Income	# of visits in a month	Average Money Spent in a month
1	500078	11,50,000	4	8,000

Thank You!

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Classification and Regression

Prof.N.L.Bhanu Murthy

Supervised Learning (Employability Prediction)

Features / Attributes / Predictors

- ✓ CGPA
- ✓ Communication Skills
- ✓ Aptitude
- ✓ Programming Skills

Response / Target

- ✓ Job Offered?

S.No.	CGPA	Communication Skills	Aptitude	Programming Skills	Job Offered?
1	9.1	Average	Good	Excellent	Yes
2	8.4	Good	Good	Good	Yes
3	8.3	Poor	Average	Average	No
4	7.1	Average	Good	Average	No
5	8.2	Good	Excellent	Excellent	No

Supervised Learning (Predicting price of a used car)

Features / Attributes / Predictors

- ✓ Brand
- ✓ Year (Mfg)
- ✓ Engine Capacity
- ✓ Mileage
- ✓ Distance travelled
- ✓ Cab?

Response / Target

- ✓ Price (in Rs.)

S.No.	Brand	Year (Mfg)	Engine Capacity	Mileage	Distance travelled	Cab?	Price (in Rs.)
1.	Honda City ZX	2008	1100	10.5	45000	N	3,50,000
2							
3							
4							

Supervised Learning

Employability Prediction

Features

- ✓ CGPA
- ✓ Communication Skills
- ✓ Aptitude
- ✓ Programming Skills

Response / Target

- ✓ Job Offered?

Predicting price of a used car

Features

- ✓ Brand
- ✓ Year (Mfg)
- ✓ Engine Capacity
- ✓ Mileage
- ✓ Distance travelled
- ✓ Cab?

Response / Target

- ✓ Price (in Rs.)

Classification and Regression

Classification problems are supervised Learning problems where target/response variables take only discrete (finite/countable) values.

Example: Employability prediction

Regression problems are supervised learning problems where target / response is a continuous variable (or equivalently can take any real number).

Example: Predicting price of a used car

Classification and Regression - Examples

Classification

- ✓ Predicting whether a patient has a particular disease or not.
- ✓ Hand written digit recognition
- ✓ Email spam detection

Regression

- ✓ Predicting house/property price
- ✓ Predicting stock market price
- ✓ Predicting sales of a product

Thank You!

Post Graduate Certificate Programme in Artificial Intelligence and Machine Learning

Prof.N.L.Bhanu Murthy

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Post Graduate Certificate Programme in AI & ML

S.No.	Course	Duration
1	Regression	5 Weeks
2	Feature Engineering	4 Weeks
3	Classification	9 Weeks
4	Unsupervised Learning & Association Rule Mining	7 Weeks
5	Text Mining	5 Weeks
6	Deep Learning & Artificial Neural Networks	6 Weeks
7	Capstone Project	8 Weeks

Duration: 44 weeks

- ✓ Refresher course in Python
- ✓ No refresher course on mathematical, statistical and probability foundations. Relevant topics will be covered as and when they are required in the course

Course 1: Regression

- Building simple and multiple regression models using
  - ✓ Gradient / Stochastic Gradient / Mini-Batch Gradient Descent Algorithm
  - ✓ Solving normal equations
- Evaluation Measures (R^2, MSE)
- Model Selection
- Overfitting
- Ridge & Lasso Regression
- Forward & Backward stepwise feature selection

Duration: 5 weeks

Course 2: Feature Engineering

- Types of data and its sources, data quality (Missing values, Noisy data)
- Data Preprocessing - Aggregation and Sampling, Feature Creation, Discretization and Banalization, Data Transformation
- Feature Subset Selection
- Dimensionality Reduction – Principal Component Analysis
- Measures of Similarity and Dissimilarity
- Visualization - Box / scatter plots, Contour plots, Heat maps, Parallel Coordinates, TSNE

Duration: 4 weeks

Course 3: Classification

- Types of classification algorithms – Discriminative models, Probabilistic Generative models and , Tree based models
- Nearest-neighbor Methods
- Naïve Bayes Classifier
- Logistic Regression
- Decision Tree
- Support Vector Machines
- Ensemble Methods

Duration: 9 weeks

Course 4:
Unsupervised Learning & Association Rule Mining

- K-Means & EM Algorithm
- Hierarchical Clustering
- Density Based Clustering
- Assessing Quality of Clustering
- Association Rule Mining
- Time series Prediction and Markov Process

Duration: 7 weeks

Course 5: Text Mining

- Document vectorization, Information Retrieval Pipeline, Stemming, Lemmatization, Wild card query using K-Gram index
- Parts of Speech Tagging
- Topic modelling using LDA
- Sentiment Analysis
- Recommender Systems – Collaborative filtering, metrics

Duration: 5 weeks

Course 6: Deep Learning and Artificial Neural Networks

- Artificial Neural Networks, Back propagation algorithm
- Sequence Modeling in Neural Network – RNN, LSTM
- Deep learning - CNN, RCNN, Faster RCNN
- Auto encoders with Deep Learning
- Generative deep learning models - Boltzmann Machine, Restricted Boltzmann Machine, Deep Belief Machines, GAN

Duration: 6 weeks

Course 7: Capstone Project

- Real life problems encompassing a typical data science pipeline
- Jointly mentored by the industry experts and faculty.
- Comparative study of the relevant techniques covered in the course.
- Fortnight review of progress of the project.

Duration: 8 weeks

Evaluation

- ✓ Every course will have assignments, quizzes, minor projects and comprehensive examination
- ✓ The distribution of marks for each of these components will be detailed in the handout of each course
- ✓ For example the evaluation scheme for Regression module is as follows:

Evaluation Component	Marks	Type
Quizzes (2)	24%	Open
Assignments/Exercises	12%	Open
Minor Projects (Evaluated twice)	24%	Open
Comprehensive Examination	40%	Closed

Evaluation

- ✓ Quizzes are online examinations and are announced at the start of the course
- ✓ Comprehensive examinations for Course 1, Course 2 and Course 3 will be conducted at the end of the Course 3
- ✓ Comprehensive examinations for Course 4, Course 5 and Course 6 will be conducted at the end of the Course 6
- ✓ Successful completion of the certificate program would require completion of all the courses with a minimum C- grade in each course

Course Administration

- ✓ The video content for a week will be uploaded on the first day (Monday) of the week
- ✓ Contact session with the instructor on the following Sunday for any clarifications
- ✓ Queries should be sent to the instructor by Friday 10PM for any clarifications to be dealt on the Sunday
- ✓ Students are encouraged to make use of discussion forum to reap benefits of collaborative learning
- ✓ Teaching Assistants will be active on discussion forums

Thank You!

Linear and Polynomial Regression

Prof.N.L.Bhanu Murthy

Regression

Regression problems are supervised learning problems where target / response is a continuous variable (or equivalently can take any real number).

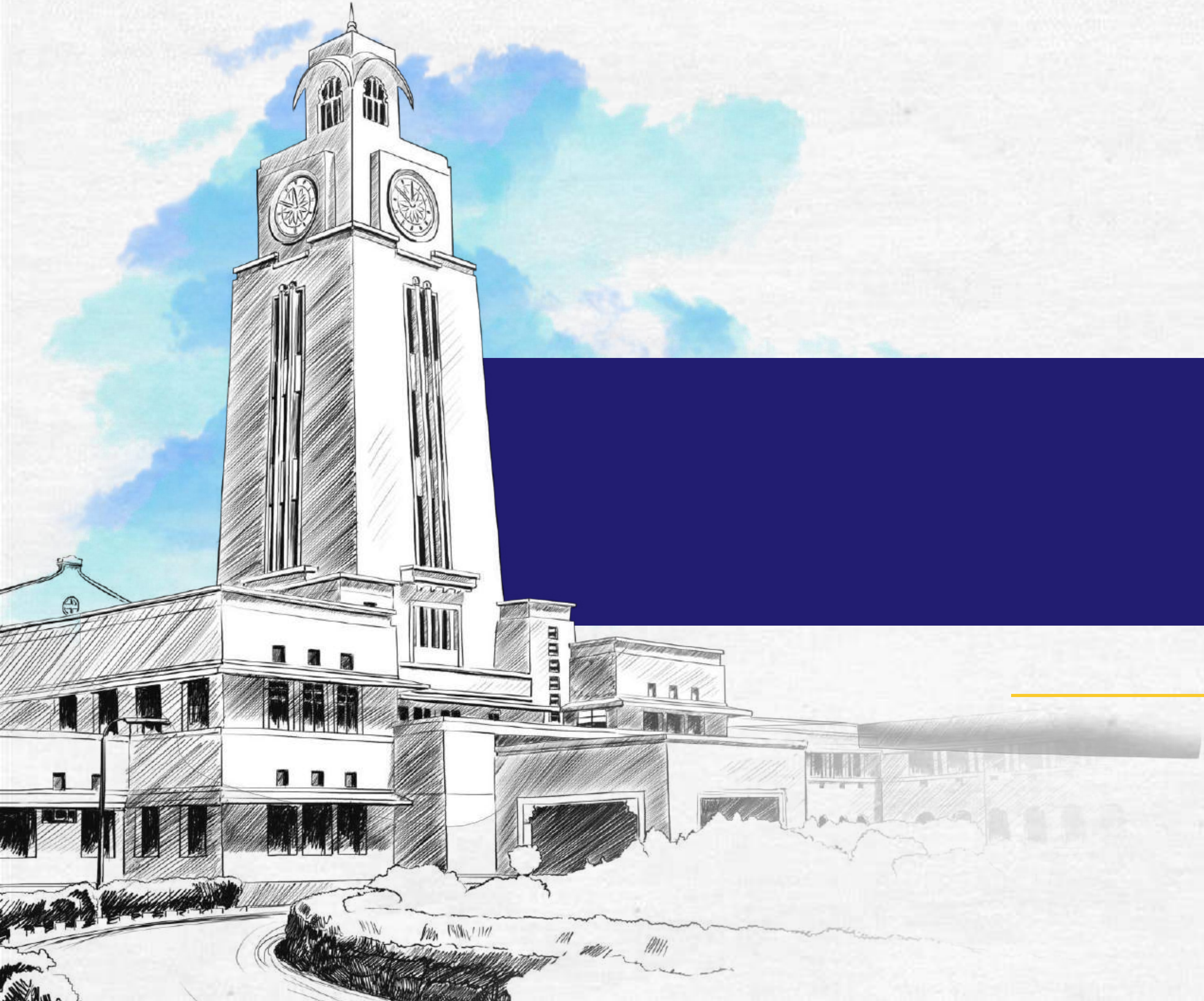
Regression

Predicting sales of an item

Advertising (in lakhs of rupees)	Sales (in lakhs of rupees)
20	625
25	730
30	850
35	1075

Regression

Thank You!



BITS Pilani

Pilani | Dubai | Goa | Hyderabad

Linear Regression

Prof.N.L.Bhanu Murthy

Linear Regression

Predicting sales of an item

Advertising (in lakhs of rupees)	Sales (in lakhs of rupees)
10	520
20	625
35	700
50	780

Linear Regression

Predicting sales of an item

Advertising (in lakhs of rupees)	Sales (in lakhs of rupees)
10	520
20	625
35	700
50	780

Linear Regression

Linear Regression

Linear Regression

Linear Regression

Linear Regression

Thank You!

Linear Regression – Convexity of Error Function

Prof.N.L.Bhanu Murthy

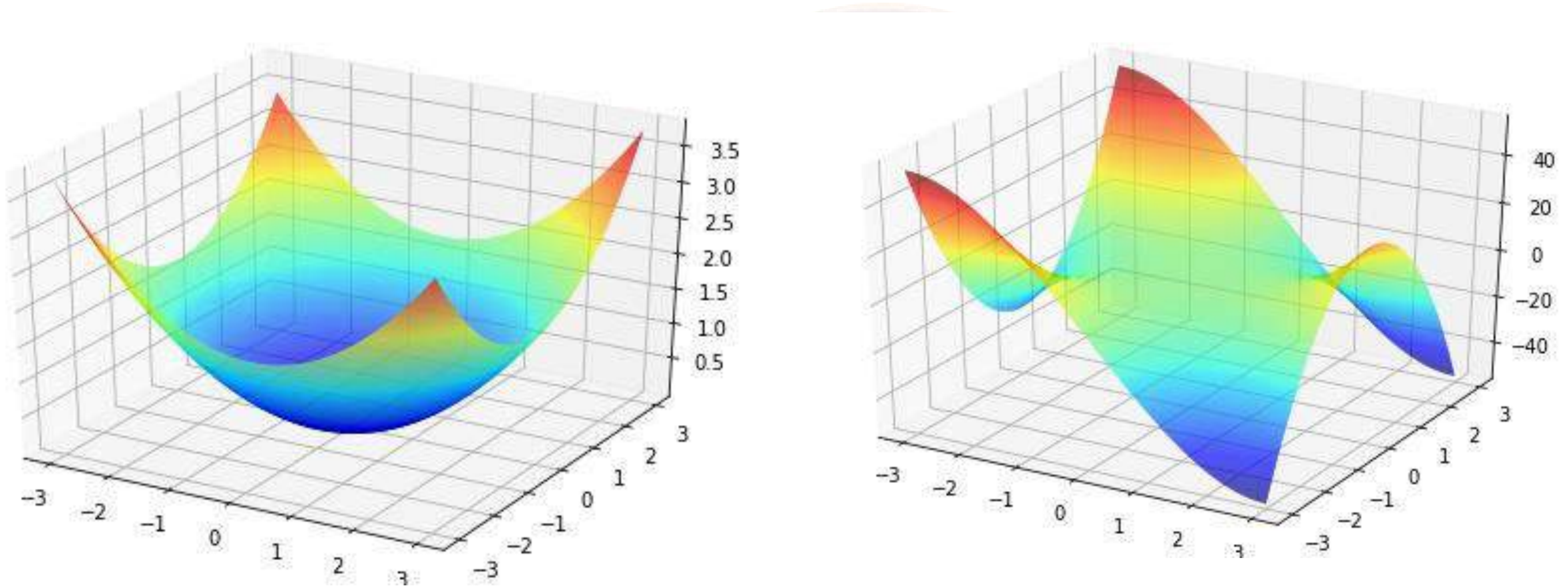
Linear Regression

Linear Regression

Predicting sales of an item

Advertising (in lakhs of rupees)	Sales (in lakhs of rupees)
10	520
20	625
35	700
50	780

Linear Regression



Linear Regression

Linear Regression

Regression

Thank You!

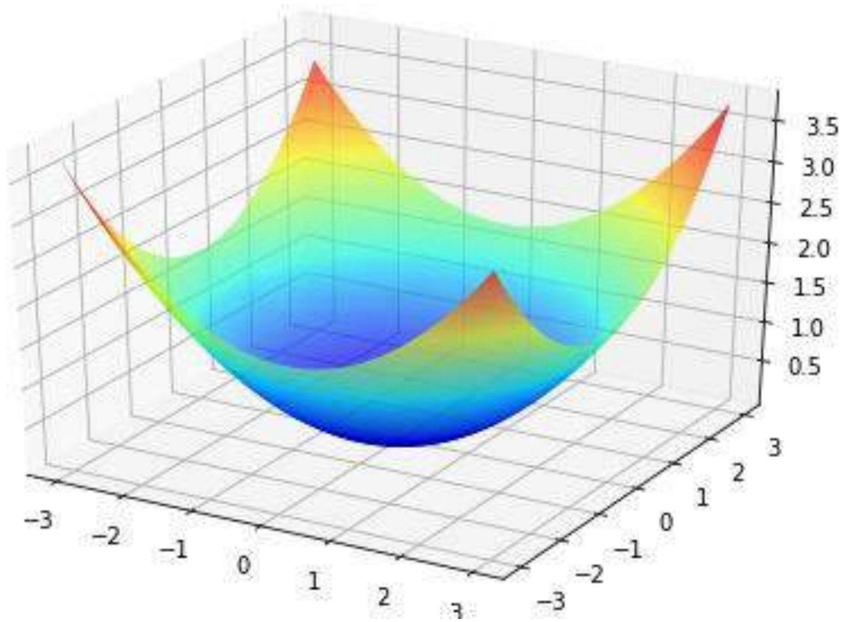
Linear Regression – Convexity of Error Function

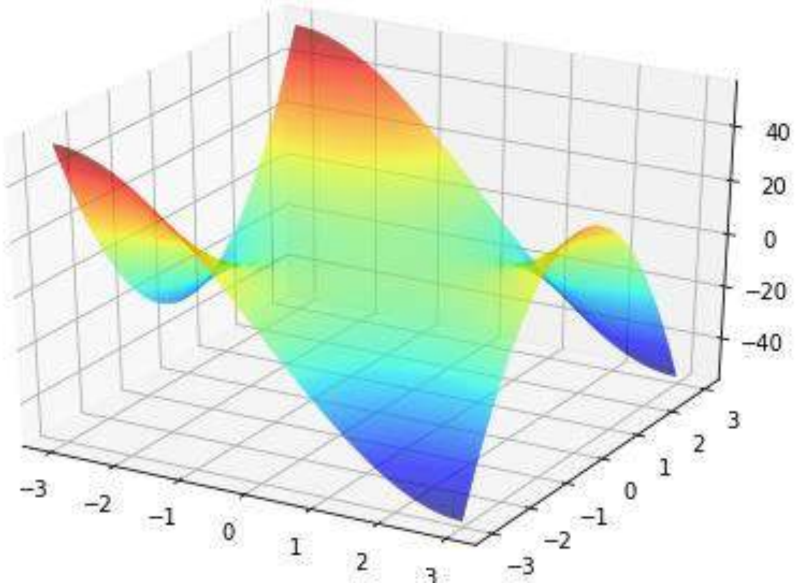
Prof.N.L.Bhanu Murthy

Linear Regression

Linear Regression

Linear Regression





Regression

Thank You!

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Linear Regression – Gradient Descent Algorithm

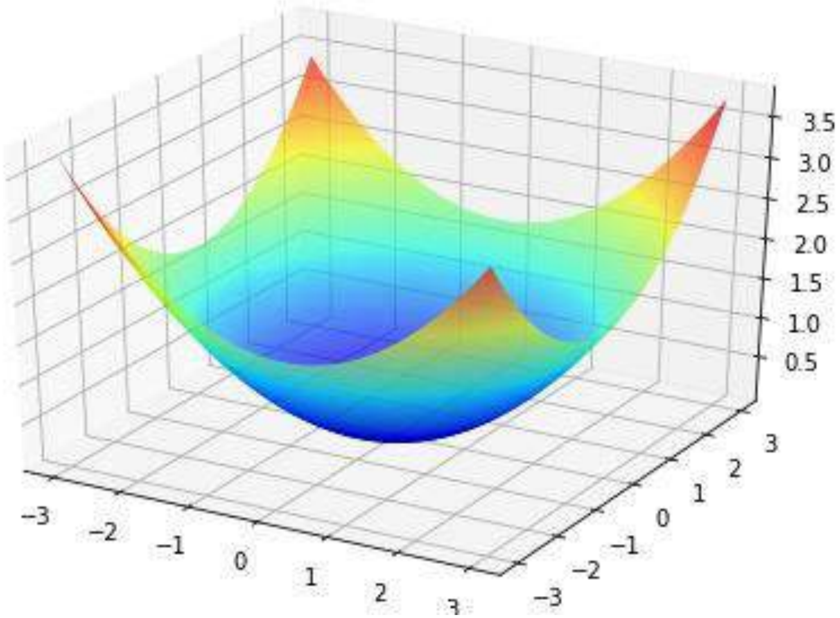
Prof.N.L.Bhanu Murthy

Linear Regression

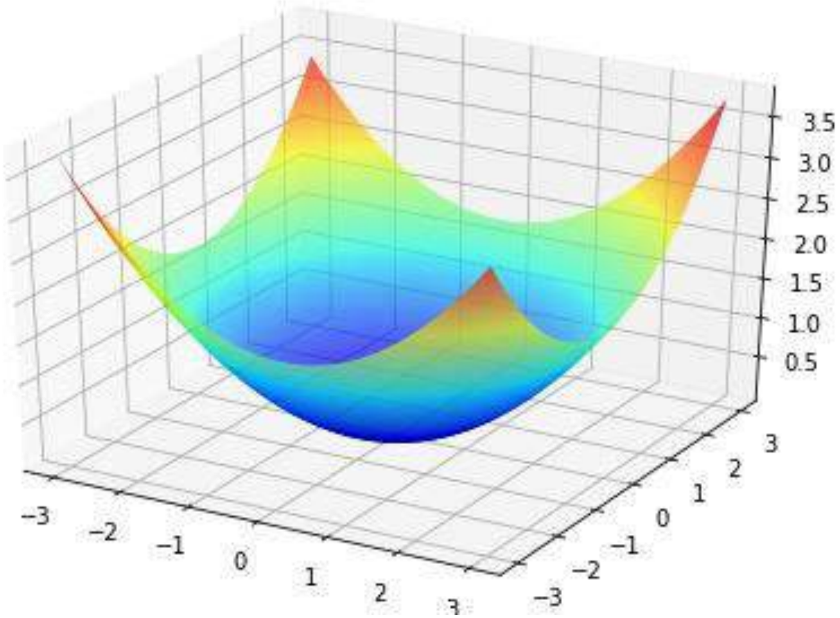
Predicting sales of an item

Advertising (in lakhs of rupees)	Sales (in lakhs of rupees)
10	520
20	625
35	700
50	780

Linear Regression



Linear Regression



Linear Regression

Linear Regression

Linear Regression

Linear Regression

Linear Regression

Linear Regression

Linear Regression

Linear Regression

Linear Regression

Linear Regression

Regression

Thank You!

Linear Regression – Gradient Descent Algorithm (2)

Prof.N.L.Bhanu Murthy

Linear Regression

Linear Regression

Linear Regression

Linear Regression

Linear Regression

Linear Regression

Linear Regression

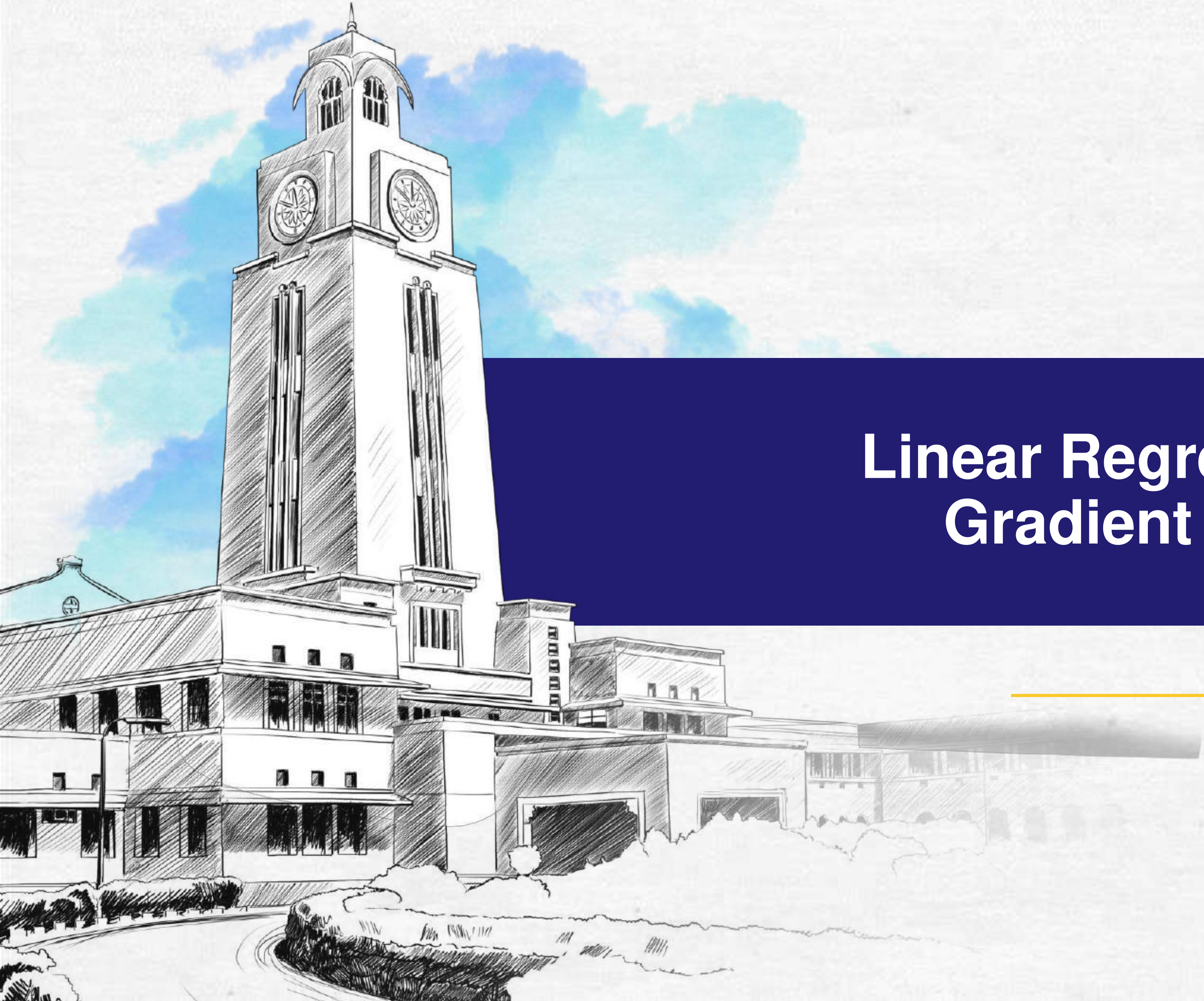
Linear Regression

Linear Regression

Linear Regression

Regression

Thank You!



BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Linear Regression – Stochastic Gradient Descent Algorithm

Prof.N.L.Bhanu Murthy

Gradient Descent Algorithm

Gradient Descent Algorithm

Gradient Descent Algorithm

Gradient Descent Algorithm

Stochastic Gradient Descent Algorithm

Mini-batch Gradient Descent Algorithm

Mini-batch Gradient Descent Algorithm

Mini-batch Gradient Descent Algorithm

Thank You!

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Discrete & Continuous Distributions

Prof.N.L.Bhanu Murthy

Discrete & Continuous Distributions

Discrete & Continuous Distributions

Random Variable

A **random variable**, usually written X, is a **variable** whose possible values are numerical outcomes of a **random** phenomenon or experiment.

Examples

- ✓ X = number of heads when the experiment is flipping a coin 20 times.
- ✓ C = the daily change in a stock price.
- ✓ R = the number of kilometers per litter you get on your car during a family vacation.

Random Variable

Discrete Random Variable

- one that takes on a ***countable*** number of values
- usually count data [Number of]
- this means you can sit down and list **all** possible outcomes without missing any

Example:

- ✓ X = sum of values on the roll of two dice: X has to be either 2, 3, 4, …, or 12.
- ✓ Y = number of accidents in Hyderabad during a week: Y has to be 0, 1, 2, 3, 4, 5, 6, 7, 8, ……………”real big number”

Random Variable

Continuous Random Variable

- one that takes on an uncountable number of values

- usually measurement data [time, weight, distance, etc]

- this means you can never list all possible outcomes even if you had an infinite amount of time

Example:

- X = time it takes you to drive home from work place: X > 0, might be 30.1 minutes measured to the nearest tenth but in reality the actual time is 30.10000001………………… minutes?)

Exercise: try to list all possible numbers between 0 and 1

Discrete Probability Disribution

A ***probability distribution (density function)*** is a table, formula, or graph that describes the values of a random variable and the probability associated with these values.

Discrete Probability Distribution

X = outcome of rolling one die

X	1	2	3	4	5	6
P(X)	1/6	1/6	1/6	1/6	1/6	1/6

Discrete Probability Notation…

- ✓ An upper-case letter will represent the ***name*** of the random variable, usually **X**.
- ✓ Its lower-case counterpart, x, will represent the ***value*** of the random variable.
- ✓ The probability that the random variable **X** will equal x is:

  P(**X** = x) or more simply P(x)
- ✓ **X** = number of heads in 10 flips of coin

  P(X = 5) = P(5) = probability of 5 heads (x) in 10 flips

Mean, Variance & Standard Deviation

- ✓ The mean of a discrete random variable is the ***weighted average*** of all of its values. The weights are the probabilities.

- ✓ This parameter is also called the expected value of X and is represented by E(X).

$$E(X) = \mu = \sum_{all\ x} xP(x)$$

- ✓ The variance is

$$V(X) = \sigma^2 = \sum_{all\ x} (x-\mu)^2 P(x)$$

- ✓ The standard deviation is

$$\sigma = \sqrt{\sigma^2}$$

Computing Mean, Variance, and Std. Dev. for Discrete Random Variable

Example A mutual fund sales person knows that there is *20% chance of closing a sale* on each call she makes. What is the ***probability distribution and mean of the number of sales*** if she plans to call three customers?

Solution:

Random Variable = X = # Sales Made in 3 Attempts

Let S denote the event of closing a sale **P(S)=.20**

Thus S^C is the event of not closing a sale, and **P(S^C)=.80**

Seems reasonable to assume that sales are **independent**.

Developing Discrete Probability Distributions

Sample Space: List of all possible outcomes

SSS : $P(X = 3) = (.2)*(.2)*(.2) = 0.008$

P(3) = .008

SSS^C : $P(X = 2) = (.2)*(.2)*(.8) = 0.032$

SS^CS : $P(X = 2) = (.2)*(.8)*(.2) = 0.032$

S^CSS : $P(X = 2) = (.8)*(.2)*(.2) = 0.032$

P(2) = .032+.032+.032
(Additive Law)

SS^CS^C : $P(X = 1) = (.2)*(.8)*(.8) = 0.128$

S^CSS^C : $P(X = 1) = (.8)*(.2)*(.8) = 0.128$

S^CS^CS : $P(X = 1) = (.8)*(.8)*(.2) = 0.128$

P(1) = .128+.128+.128
(Additive Law)

$S^CS^CS^C$: $P(X = 0) = (.8)*(.8)*(.8) = 0.512$

P(0) = .512

X	0	1	2	3
P(x)	0.512	0.384	0.096	0.008

Computing Mean for Discrete Random Variable

X	0	1	2	3
P(x)	0.512	0.384	0.096	0.008

✓ Mean = 0*(.512) + 1*(.384) + 2*(.096) + 3*(.008)

= 0 + 0.384 + 0.192 + 0.024

= 0.6

Thank You!

Discrete & Continuous Distributions (2)

Prof.N.L.Bhanu Murthy

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Continuous Probability Distributions

A random variable is continuous if it can assume any value in some interval of real numbers.

Def: Let X be a continuous random variable. A function f such that

1. $f(x) \geq 0$ for real x

2.

3.

is called probability density function.

Normal Distribution

A random variable X with probability density function

$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}$$

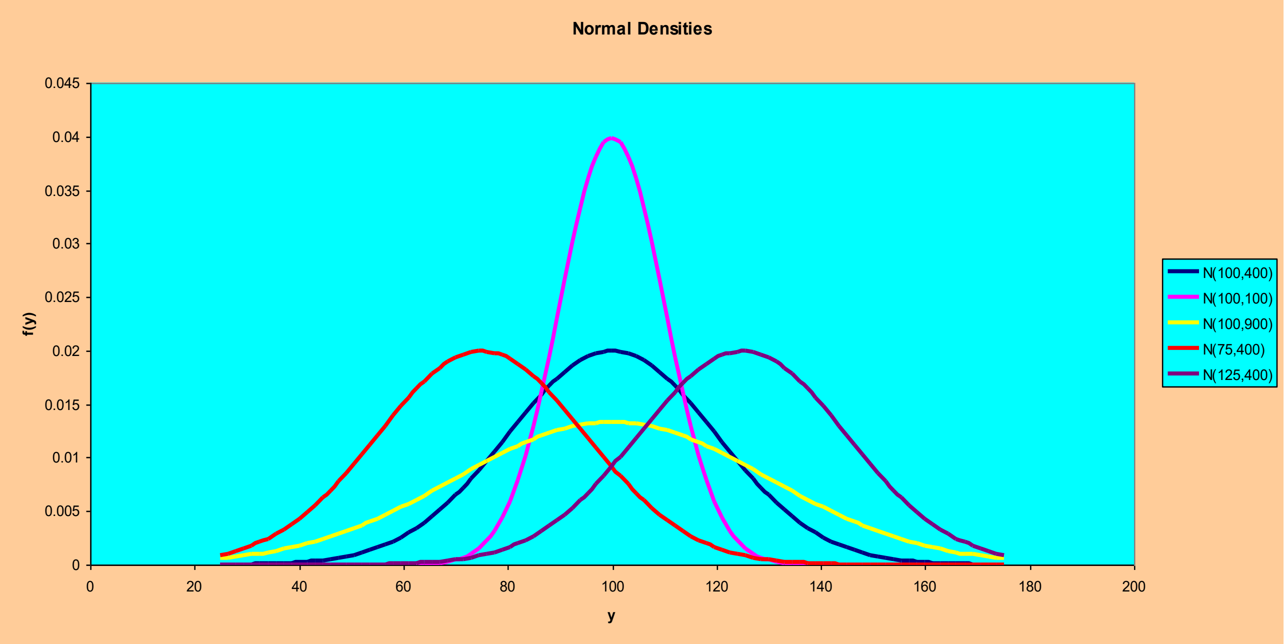
is said to have normal distribution with parameters μ and sigma.

Normal Distributions

$$f(x)=\frac{1}{\boxed{\sigma}\sqrt{2\pi}}\cdot e^{-\frac{1}{2}\left(\frac{x-\boxed{\mu}}{\sigma}\right)^2}$$

This is a bell shaped curve with different centers and spreads depending on μ and σ

Note constants:
π=3.14159
e=2.71828



Normal Distributions

Normal distribution is defined by its mean and standard deviation!!

E(X)=μ = $$\int_{-\infty}^{+\infty} x \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx$$

Var(X)=σ² = $$(\int_{-\infty}^{+\infty} x^2 \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx) - \mu^2$$

Standard Deviation(X)=σ

Normal Distributions

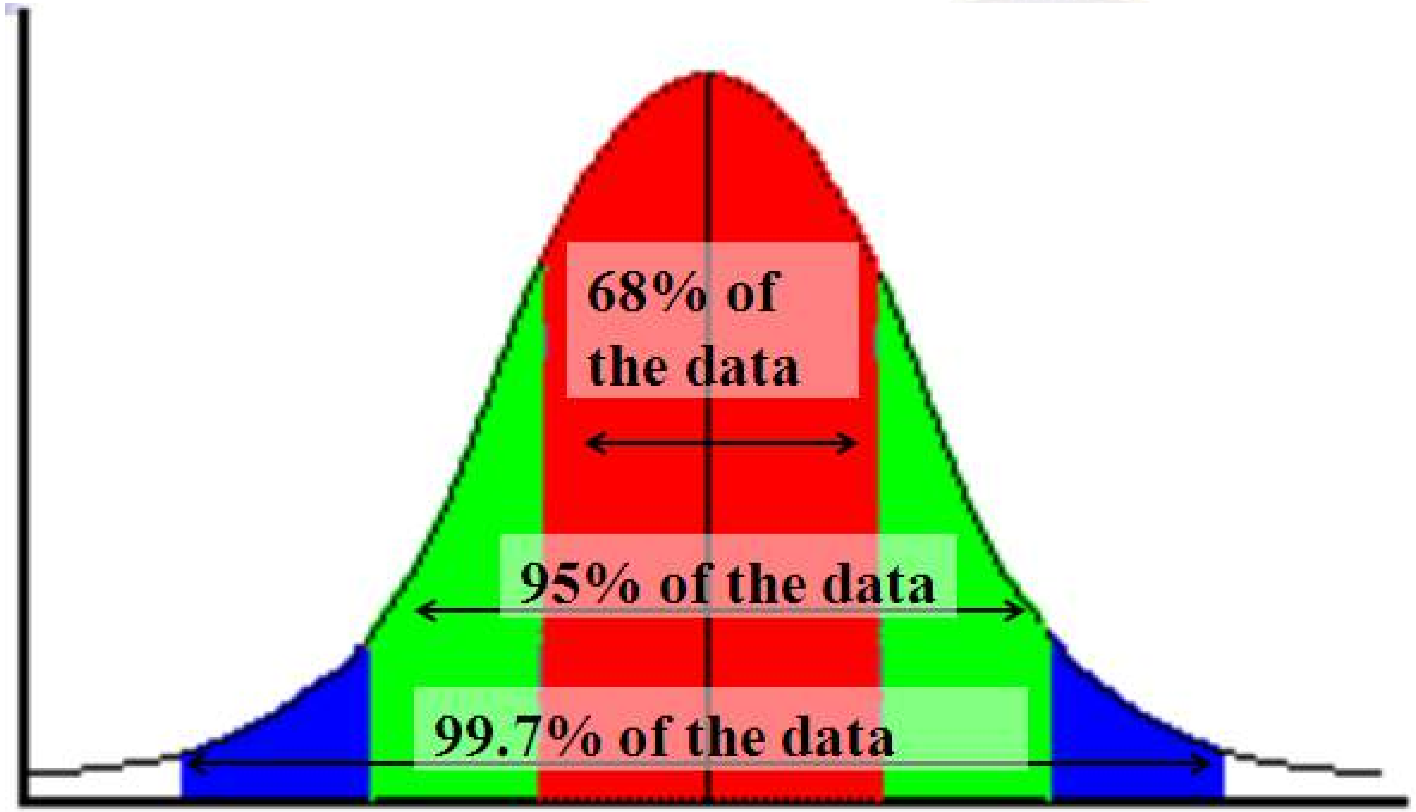
Normal Distributions

**The beauty of the normal curve:

No matter what μ and σ are,
the area between $\mu-\sigma$ and $\mu+\sigma$ is about 68%;
the area between $\mu-2\sigma$ and $\mu+2\sigma$ is about 95%; and
the area between $\mu-3\sigma$ and $\mu+3\sigma$ is about 99.7%.
Almost all values fall within 3 standard deviations.

Normal Distributions

68-95-99.7 Rule



Normal Distributions

68-95-99.7 Rule in Math terms…

$$\int_{\mu-\sigma}^{\mu+\sigma} \frac{1}{\sigma\sqrt{2\pi}} \bullet e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx = .68$$

$$\int_{\mu-2\sigma}^{\mu+2\sigma} \frac{1}{\sigma\sqrt{2\pi}} \bullet e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx = .95$$

$$\int_{\mu-3\sigma}^{\mu+3\sigma} \frac{1}{\sigma\sqrt{2\pi}} \bullet e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx = .997$$

Normal Distributions

t Distribution

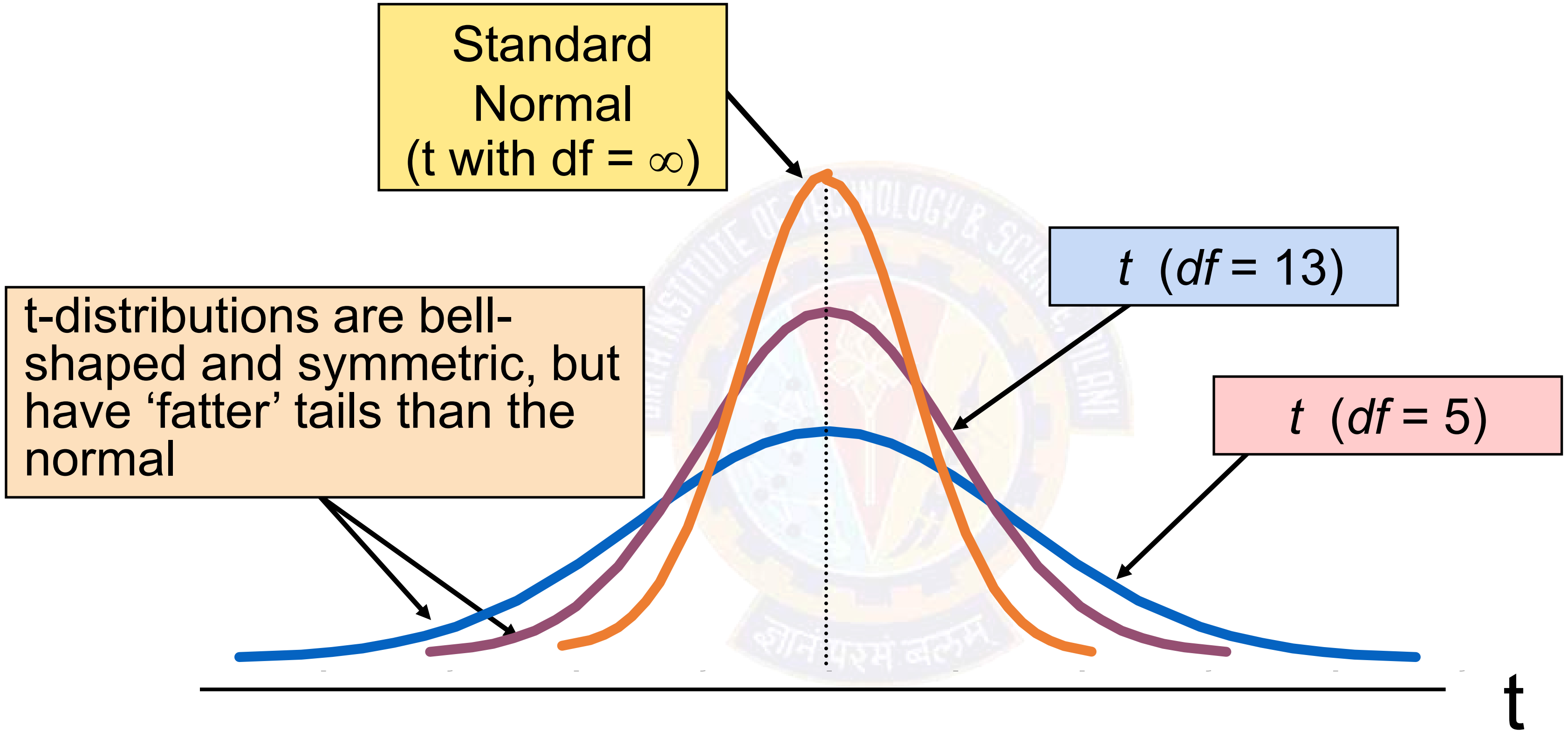
A random variable T with probability density function

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

is said to have a t distribution with v degrees of freedom and Γ (gamma) is the Gamma function defined by

$$\Gamma(x) \equiv \int_0^\infty u^{x-1}e^{-u}\,\mathrm{d}u.$$

t Distribution

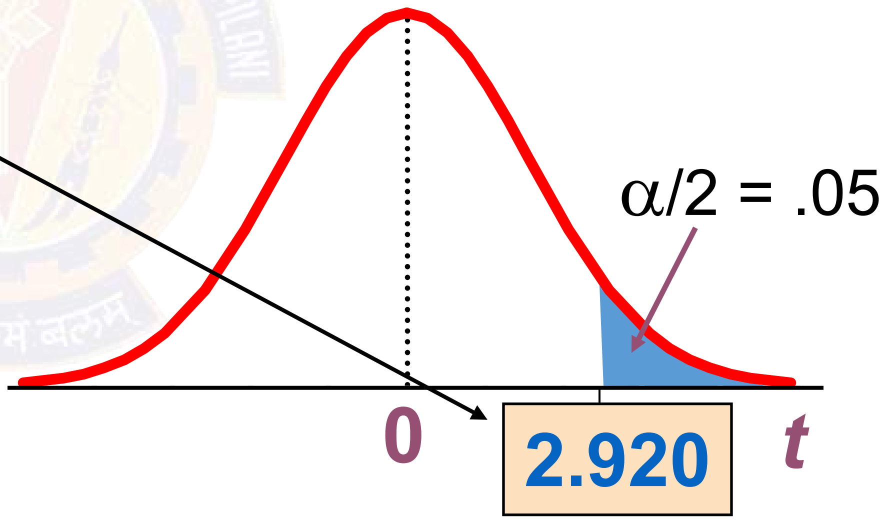


t Distribution

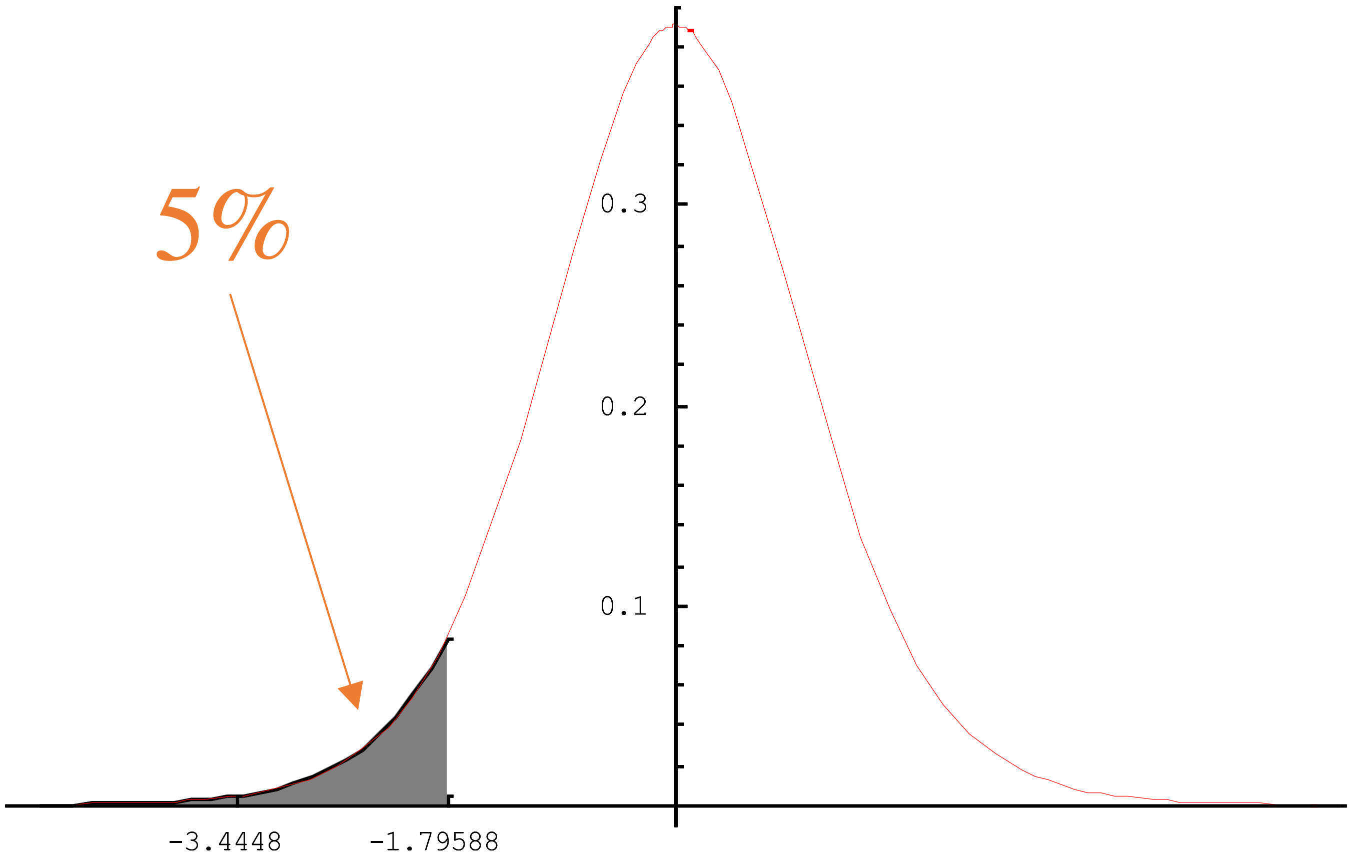
df	Upper Tail Area .25	.10	.05
1	1.000	3.078	6.314
2	0.817	1.886	2.920
3	0.765	1.638	2.353

The body of the table contains t values, not probabilities

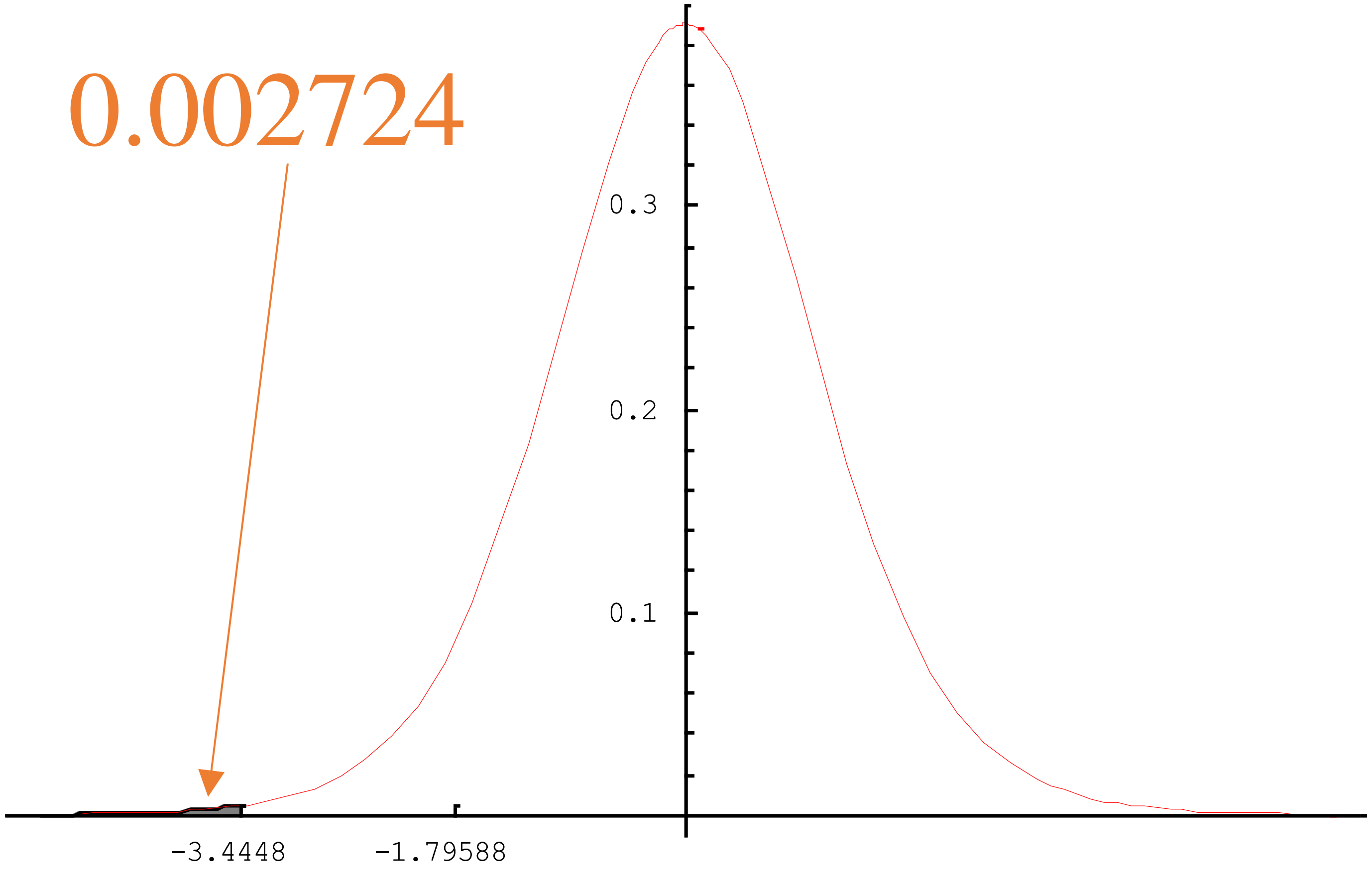
Let: n = 3
df = n - 1 = 2
α = .10
$\alpha/2$ =.05



t Distribution

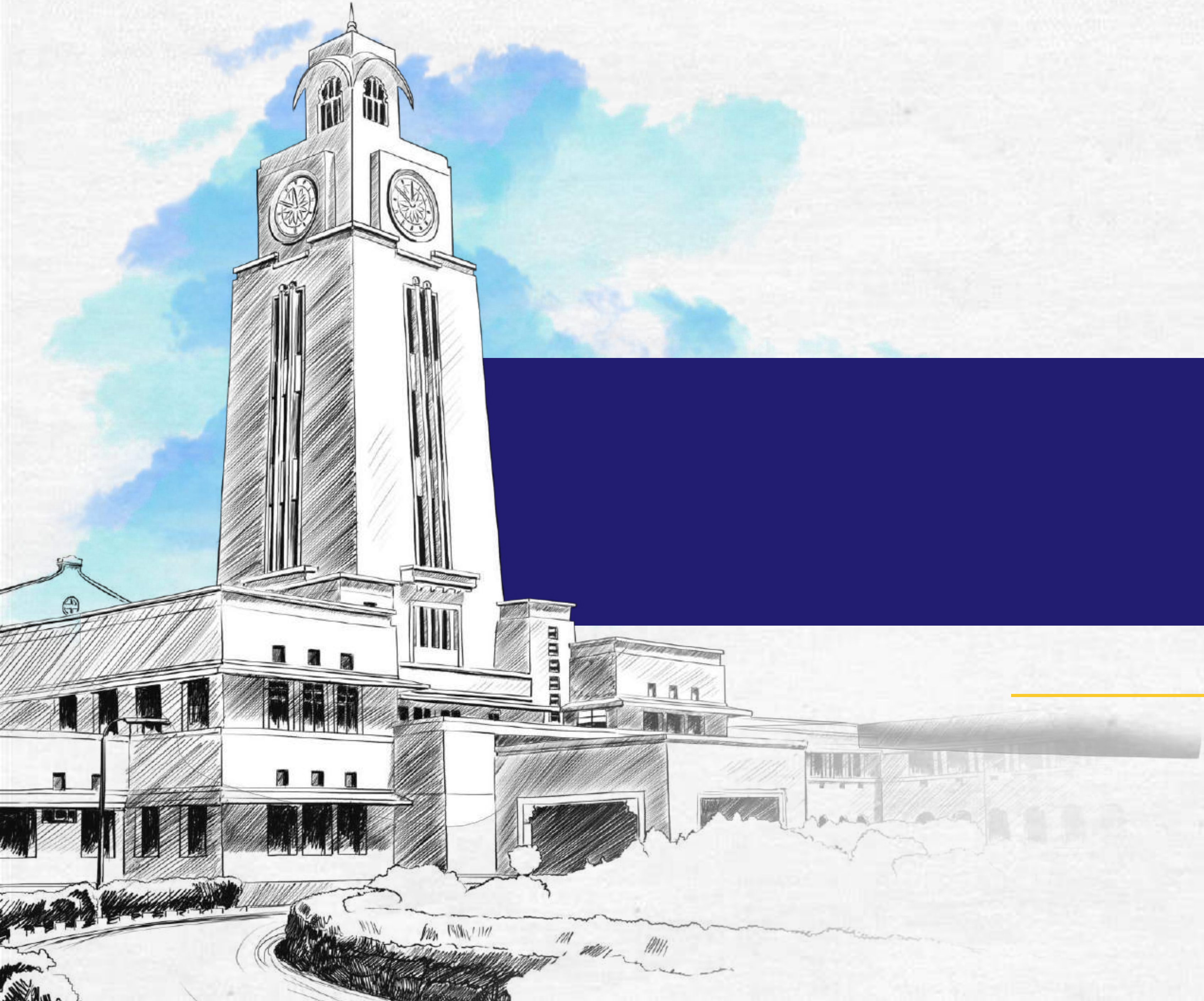


t Distribution



Continuous Probability Distributions

Thank You!



BITS Pilani
Pilani | Dubai | Goa | Hyderabad

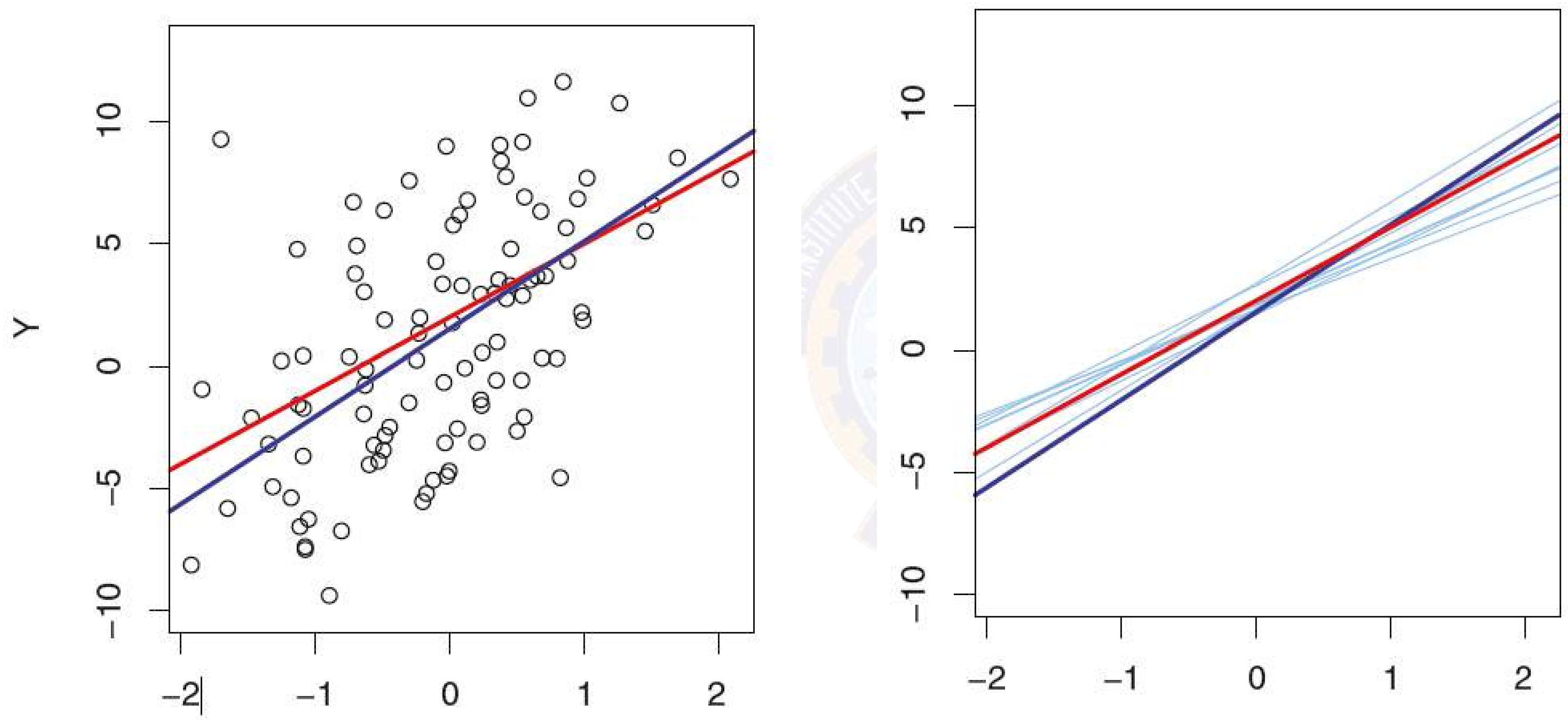
Linear Regression

Prof.N.L.Bhanu Murthy

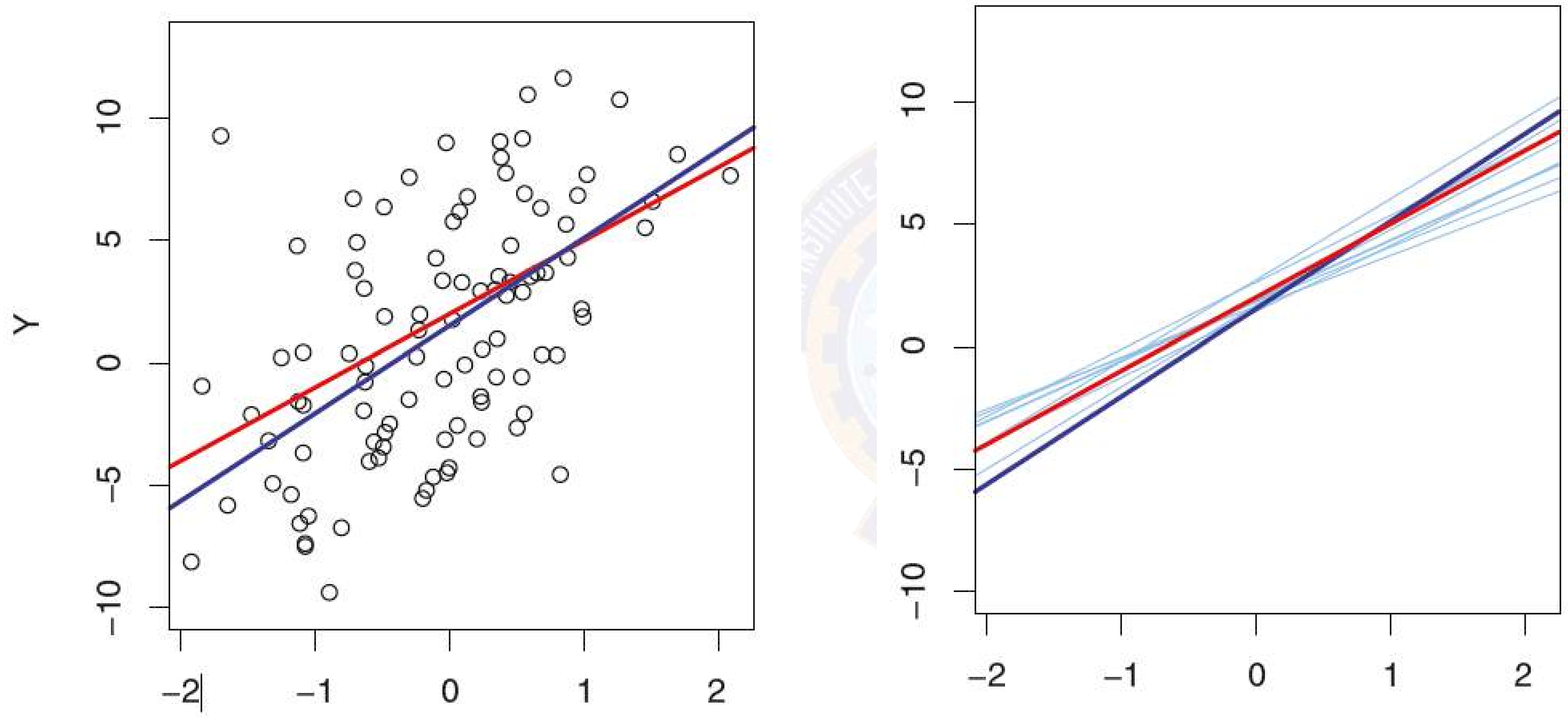
Unbiased Estimators

Unbiased Estimators

Unbiased Estimators



Unbiased Estimators



Unbiased Estimators

Unbiased Estimators

Thank You!

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Relationship between dependent and independent variable

Prof.N.L.Bhanu Murthy

Relationship between dependent and independent variable

Relationship between dependent and independent variable

Relationship between dependent and independent variable

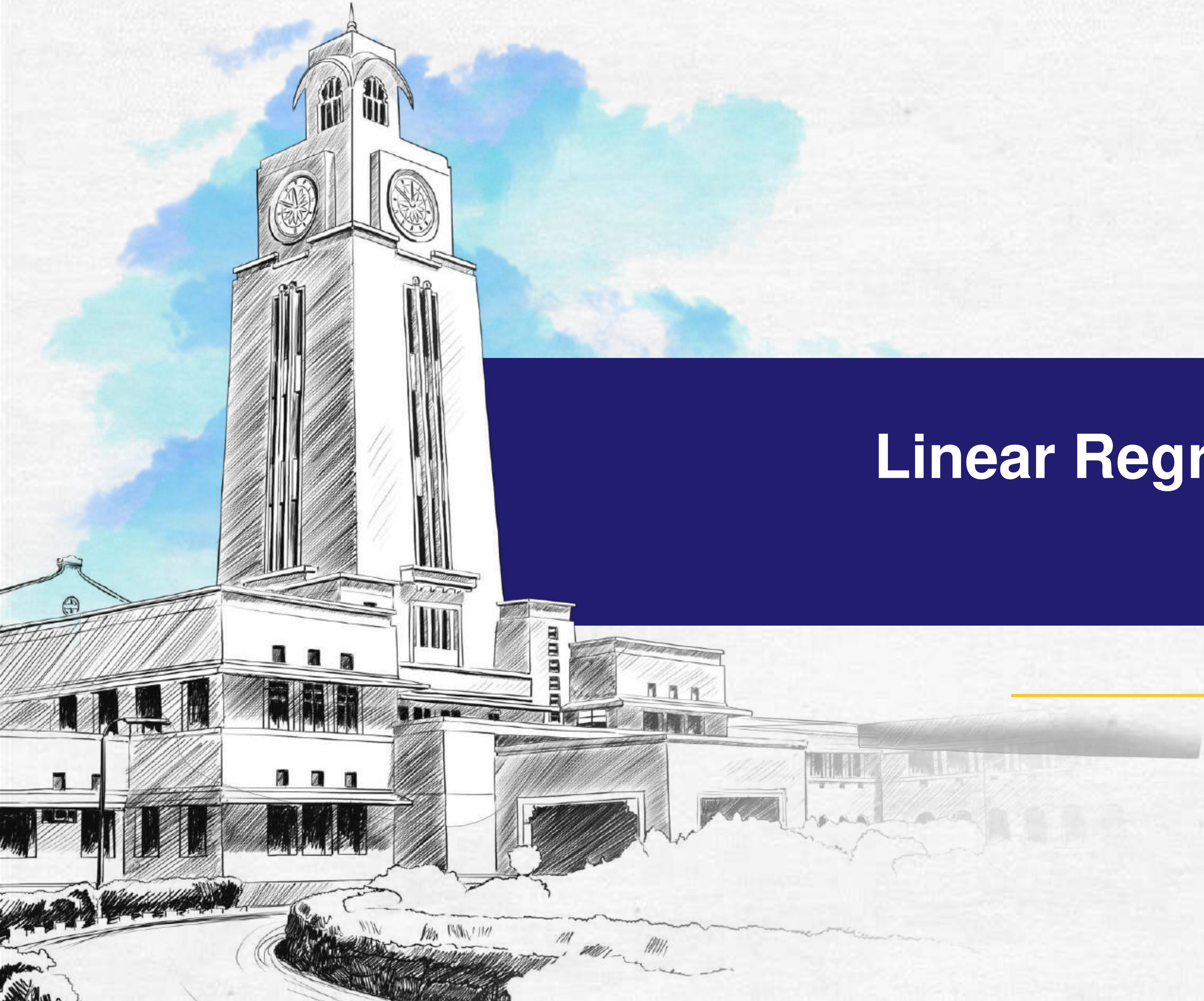
Relationship between dependent and independent variable

Relationship between dependent and independent variable

Relationship between dependent and independent variable

Relationship between dependent and independent variable

Thank You!



Linear Regression - Evaluation Measures (RSE)

Prof.N.L.Bhanu Murthy

Linear Regression

Linear Regression

Linear Regression

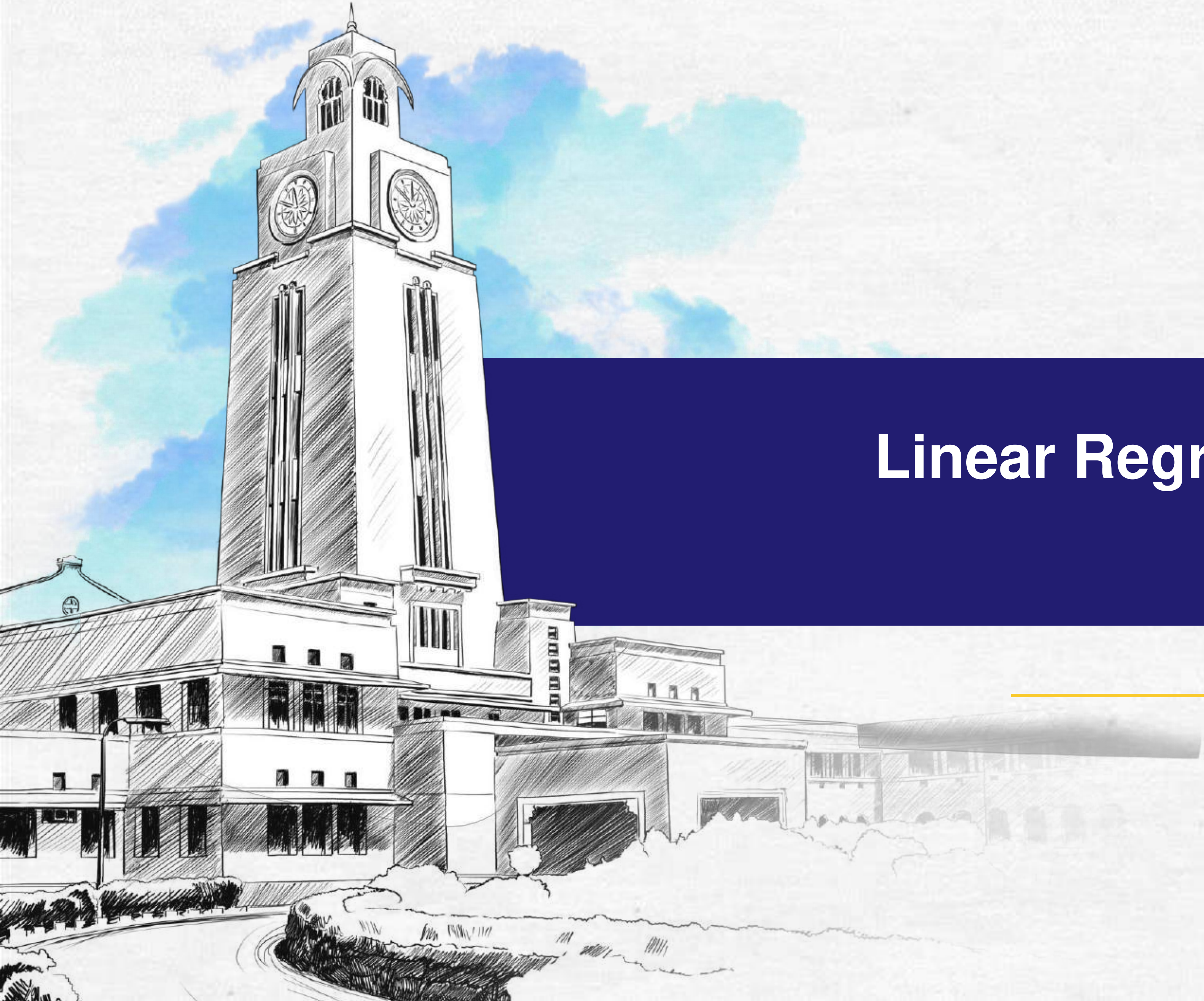
Linear Regression

	Coefficient	Std. error	t-statistic	p value
Intercept	7.15	0.56	12.76	<0.01
Advt.	1.95	0.12	16.25	<0.01

Linear Regression

Linear Regression

Thank You!



Linear Regression - Evaluation Measures (R^2)

Prof.N.L.Bhanu Murthy

Linear Regression

Predicting sales of an item

Advertising (in lakhs of rupees)	Sales (in lakhs of rupees)
10	520
20	625
35	700
50	780
20	605

Linear Regression

Predicting sales of an item

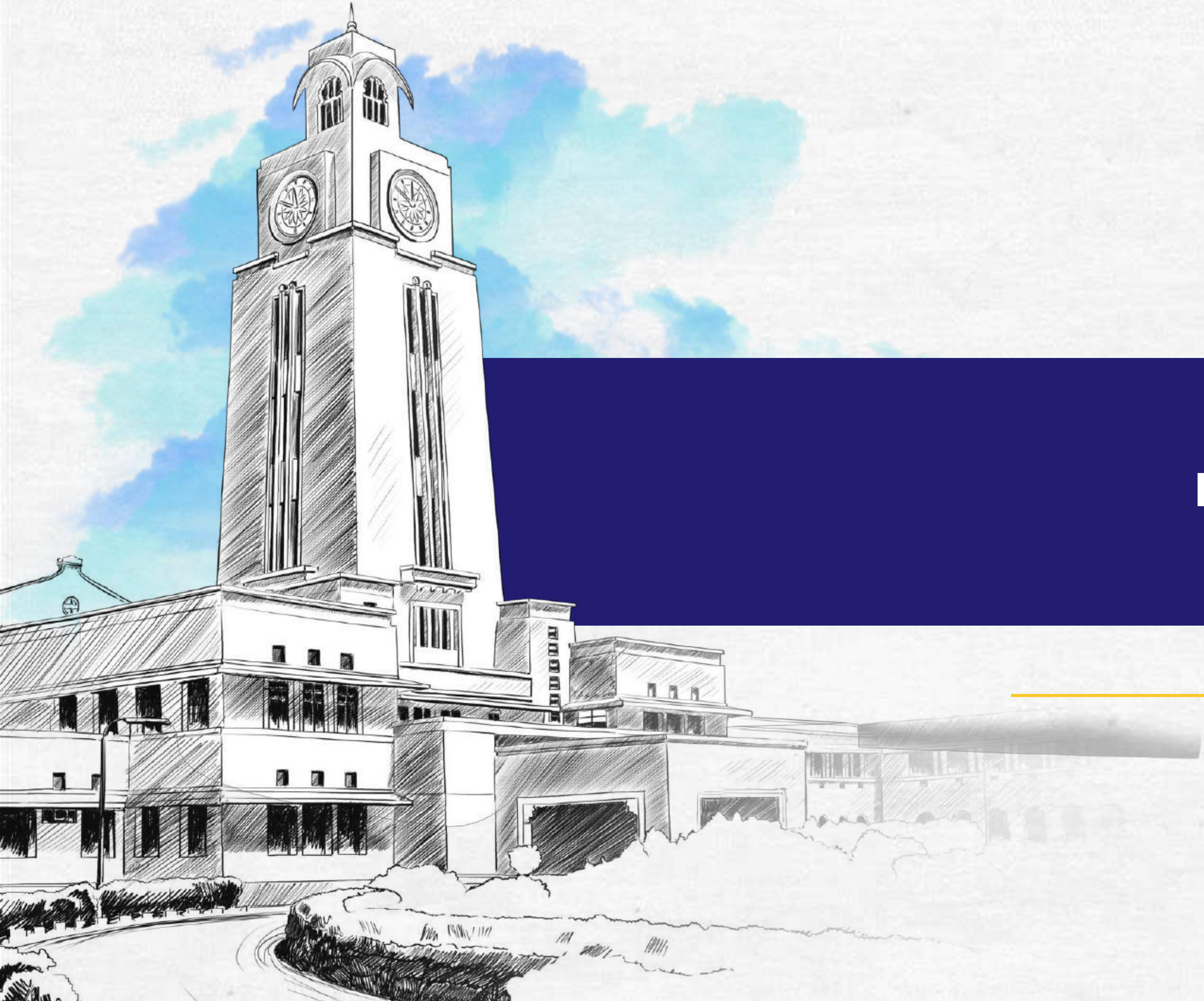
Advertising (in lakhs of rupees)	Sales (in lakhs of rupees)
10	520
20	625
35	700
50	780
20	605

Linear Regression

Predicting sales of an item

Advertising (in lakhs of rupees)	Sales (in lakhs of rupees)
10	520
20	625
35	700
50	780
20	605

Thank You!



BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Mutiple Linear Regression

Prof.N.L.Bhanu Murthy

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

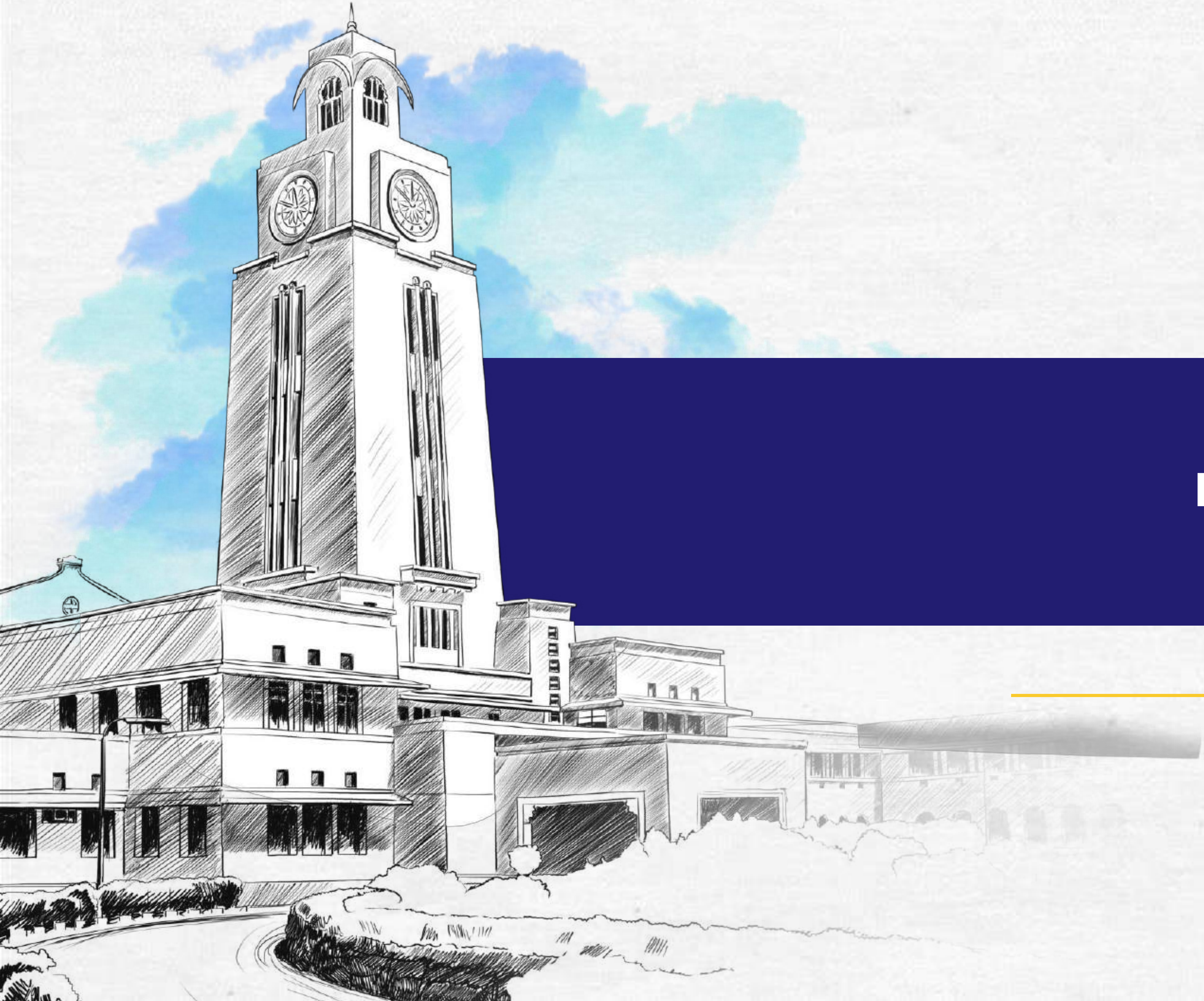
Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Thank You!



BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Mutiple Linear Regression

Prof.N.L.Bhanu Murthy

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

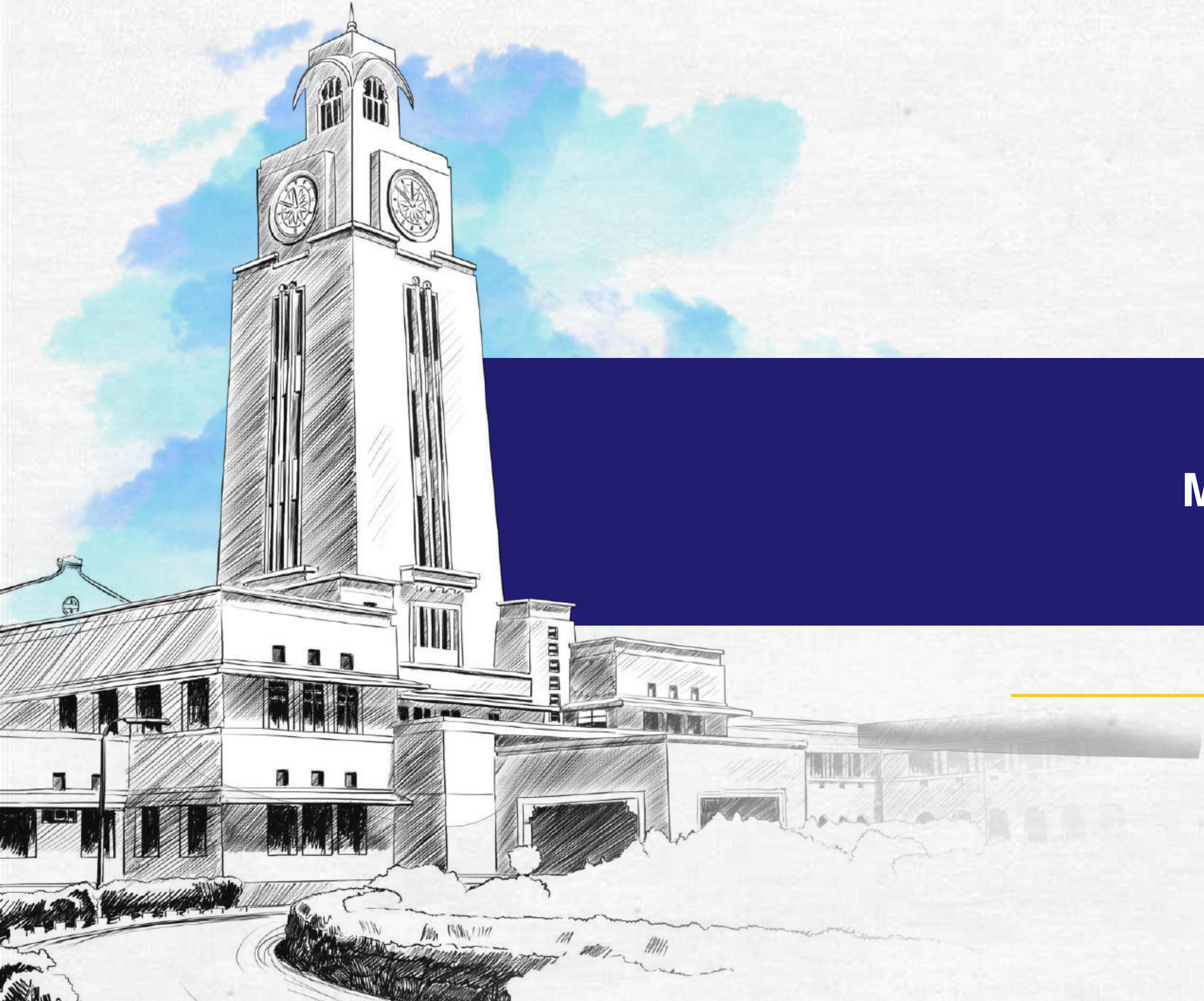
Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Thank You!



BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Multiple Linear Regression

Prof.N.L.Bhanu Murthy

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

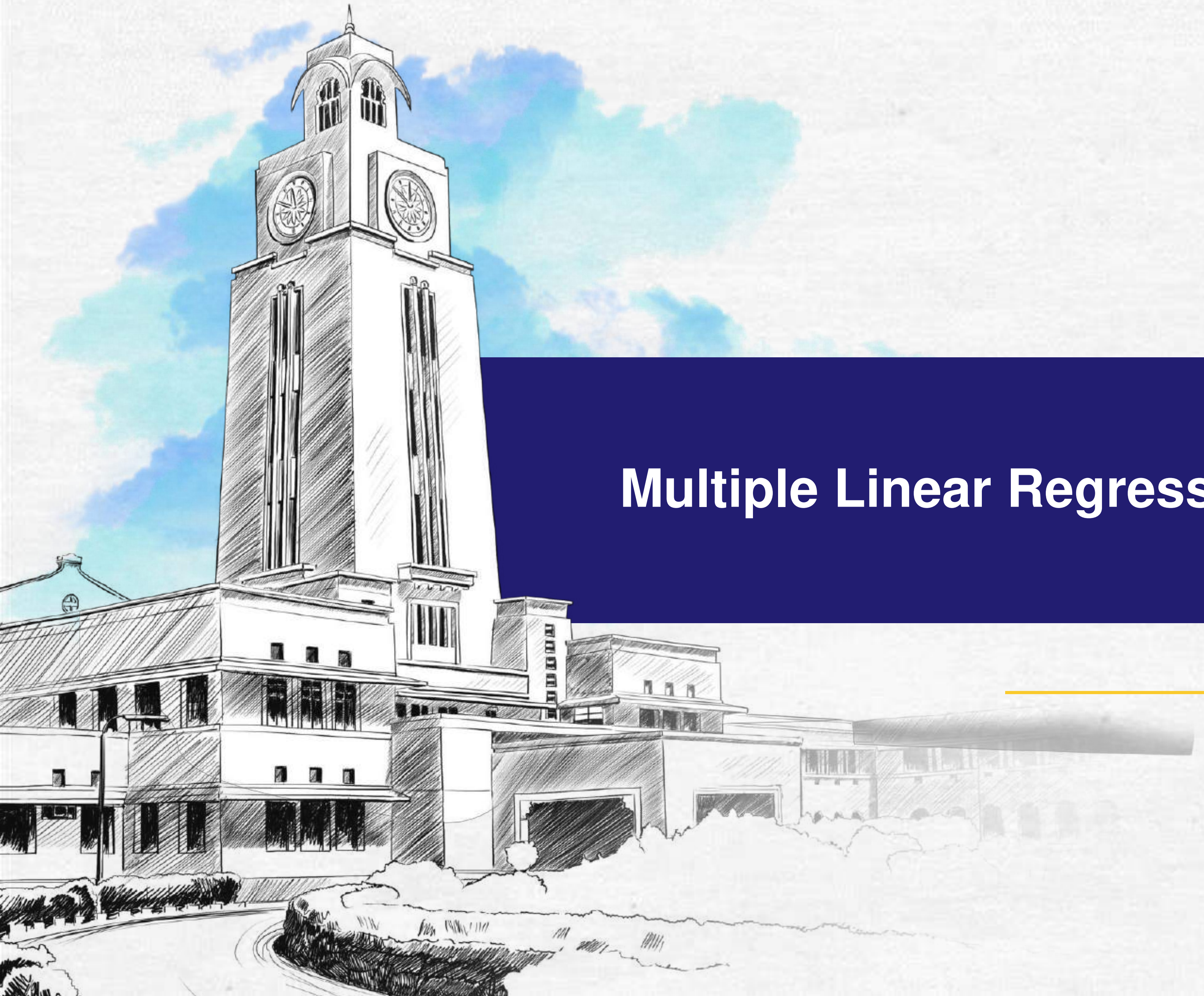
Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Multiple Linear Regression

Thank You!



BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Multiple Linear Regression: Subset Selection

Prof.N.L.Bhanu Murthy

Best Subset Selection

Best Subset Selection

Training Dataset – 2/3rd of Data Set, Testing Dataset – 1/3rd of Data Set

1. Let M_0 denote the *null model* , which contains no predictors. This model simply predicts the sample mean for each observation.

2. For $k = 1, 2, . . . D$:

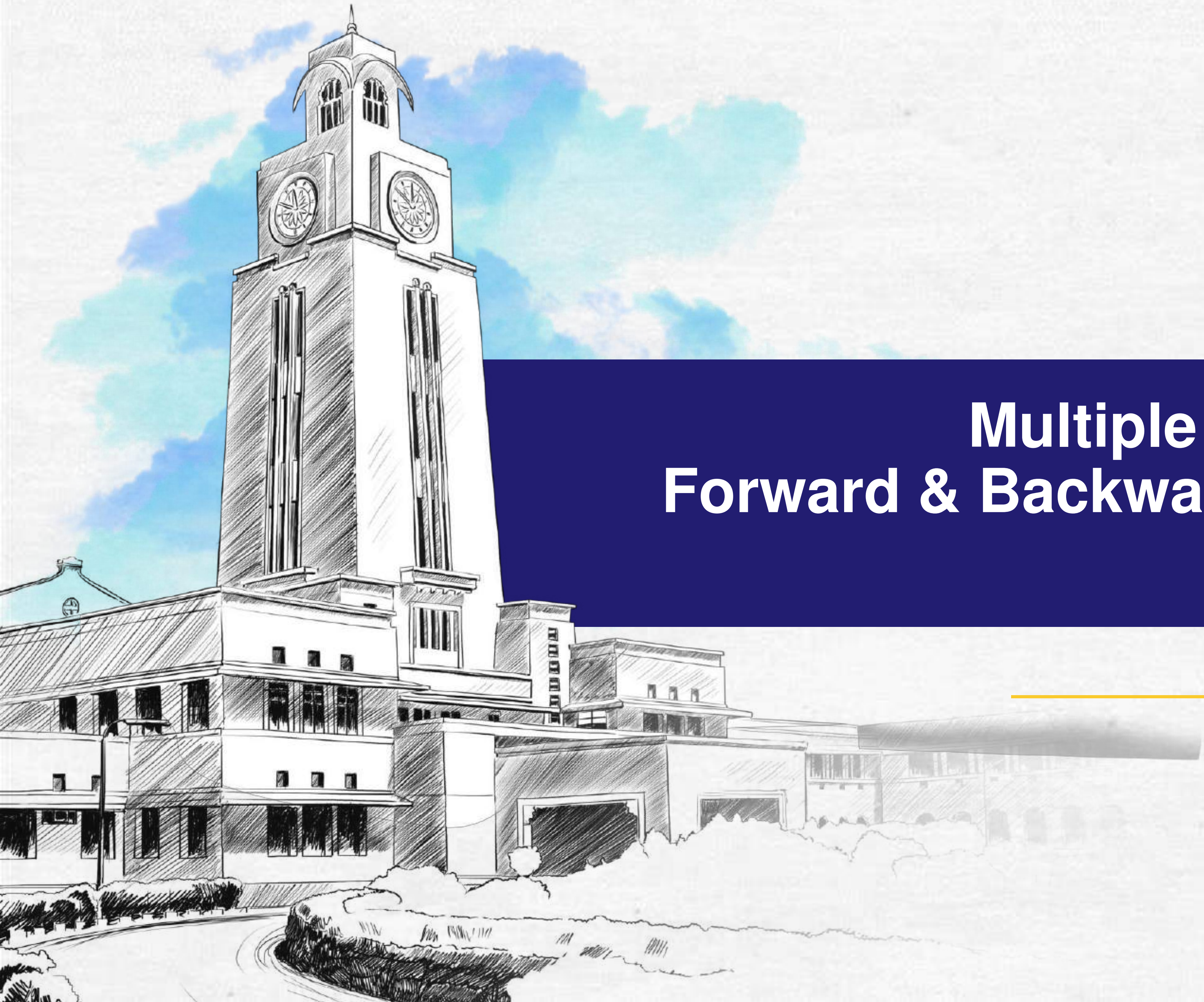
(a) Fit all $\binom{D}{k}$ models that contain exactly k predictors on training dataset.

(b) Pick the best among these $\binom{D}{k}$ models, and call it M_k. Here *best* is defined as having the smallest RSS on training dataset (or equivalently largest R^2).

3. Select a single best model from among $M_0, . . . , M_D$ having the smallest RSS on testing error (or equivalently largest R^2).

Best Subset Selection

Thank You!



Multiple Linear Regression: Forward & Backward Stepwise Subset Selection

Prof.N.L.Bhanu Murthy

Subset Selection

Forward stepwise Selection

Training Dataset – 2/3rd of Data Set, Testing Dataset – 1/3rd of Data Set

1. Let M_0 denote the *null model* , which contains no predictors. This model simply predicts the sample mean for each observation.

2. For $k = 0, 1, 2, . . . D-1$:
(a) Consider all $D - k$ models that augment the predictors in M_k with one additional predictor.
(b) Choose the *best* among these $D - k$ models, and call it M_{k+1}. Here *best* is defined as having smallest RSS on training dataset (or highest R^2).

3. Select a single best model from among $M_0, M_1, . . . , M_D$ having the smallest RSS on testing error (or equivalently largest R^2).

Forward stepwise Selection

Backward stepwise Selection

Training Dataset – 2/3rd of Data Set, Testing Dataset – 1/3rd of Data Set

1. Let M_D denote the *full* model, which contains all p predictors.

2. For $k = D, D - 1, \ldots, 1$:
(a) Consider all k models that contain all but one of the predictors in M_k, for a total of $k - 1$ predictors.
(b) Choose the *best* among these k models, and call it M_{k-1}. Here
best is defined as having smallest RSS on training dataset (or highest R^2).

3. Select a single best model from among $M_0, M_1, \ldots, M_D$ having smallest RSS on testing error (or equivalently largest R^2).

Backward stepwise Selection

Lasso Regression

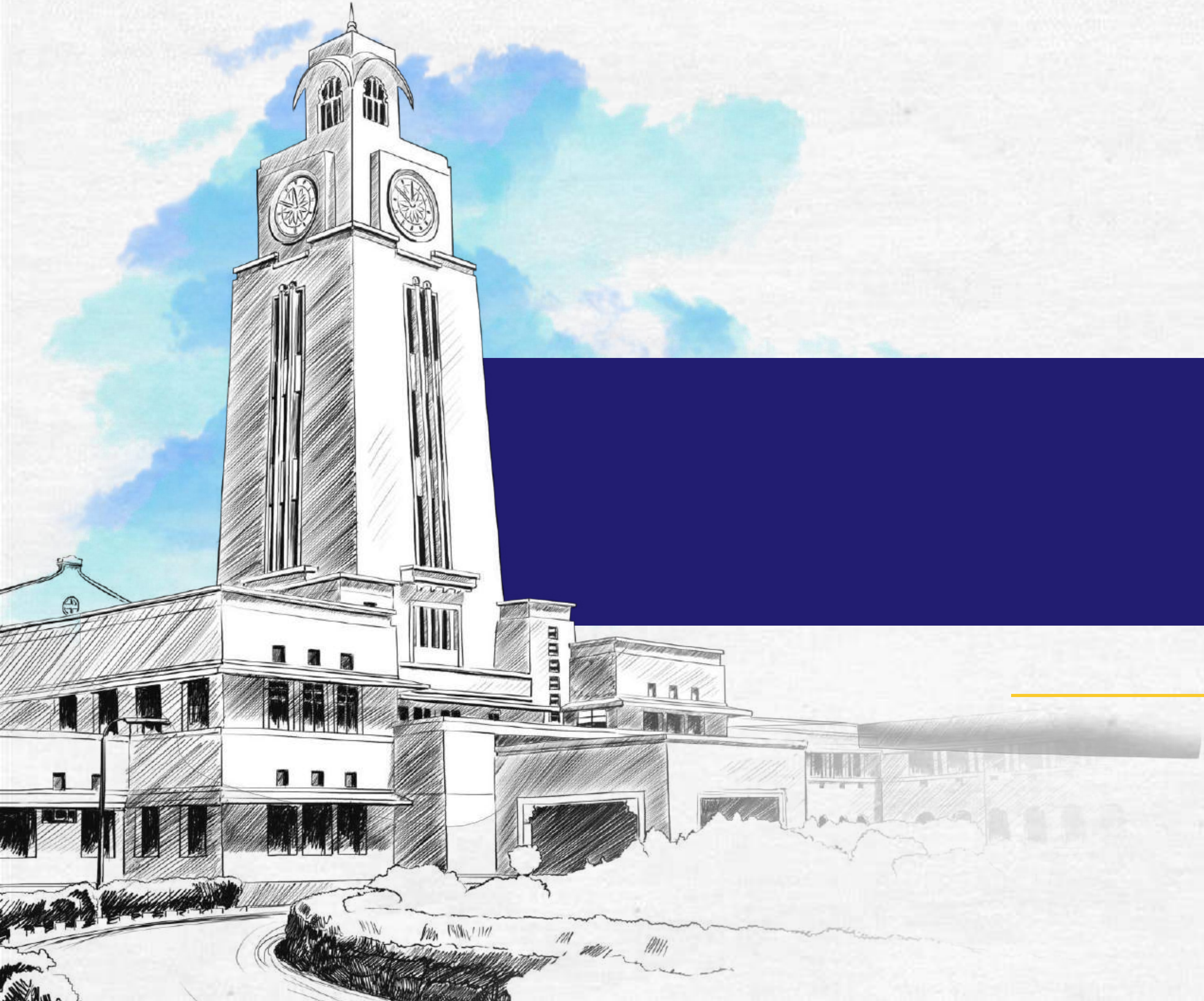
Lasso Regression

Lasso Regression

Lasso Regression

Lasso Regression

Thank You!



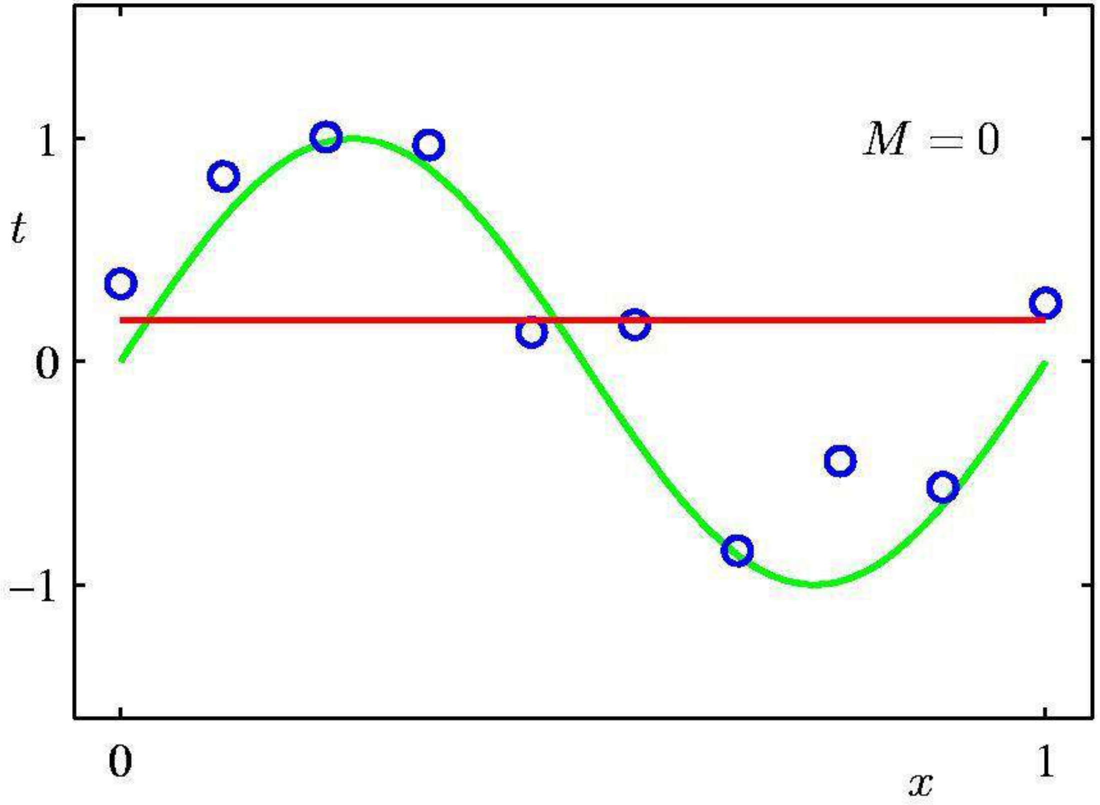
BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Model Selection

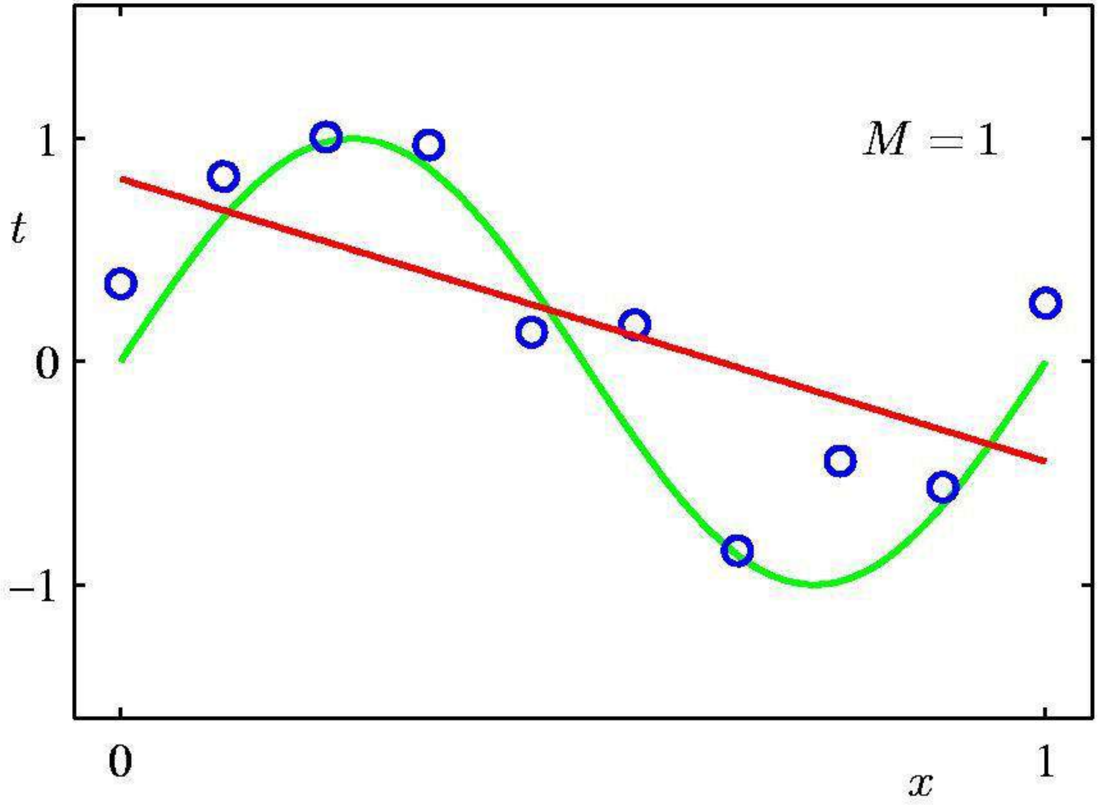
Prof.N.L.Bhanu Murthy

Model Selection

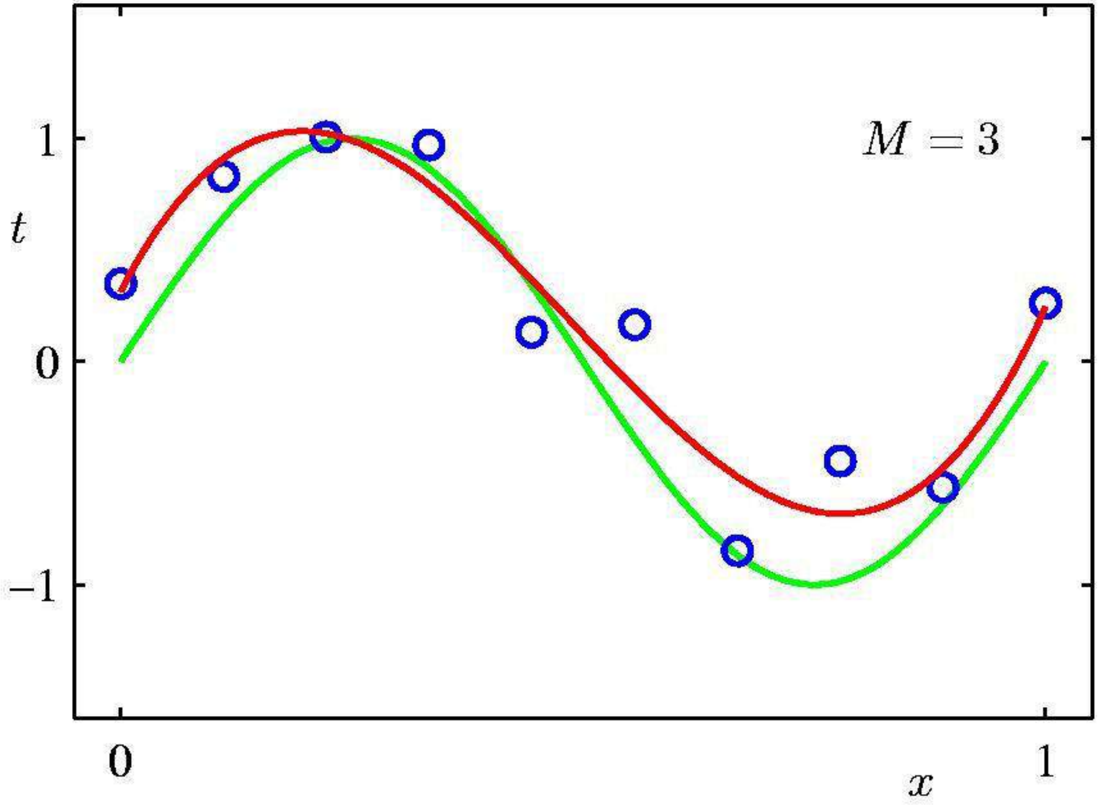
Model Selection - 0^{th} Order Polynomial



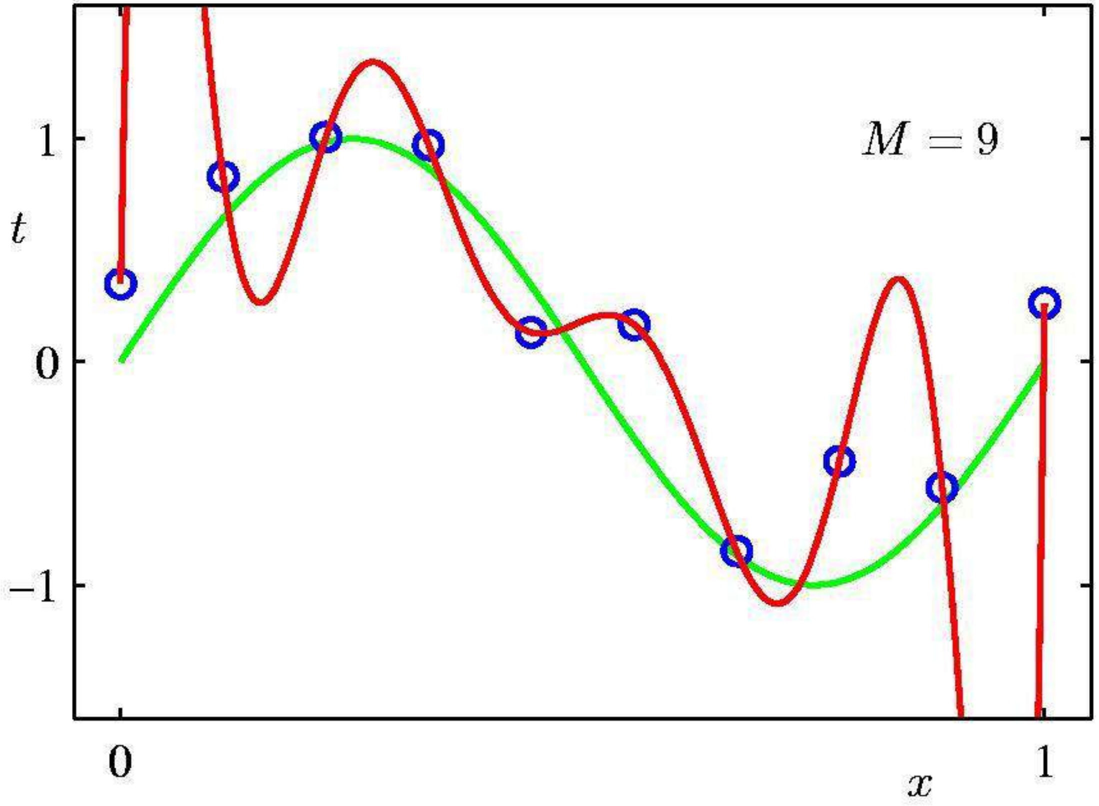
Model Selection - 1st Order Polynomial



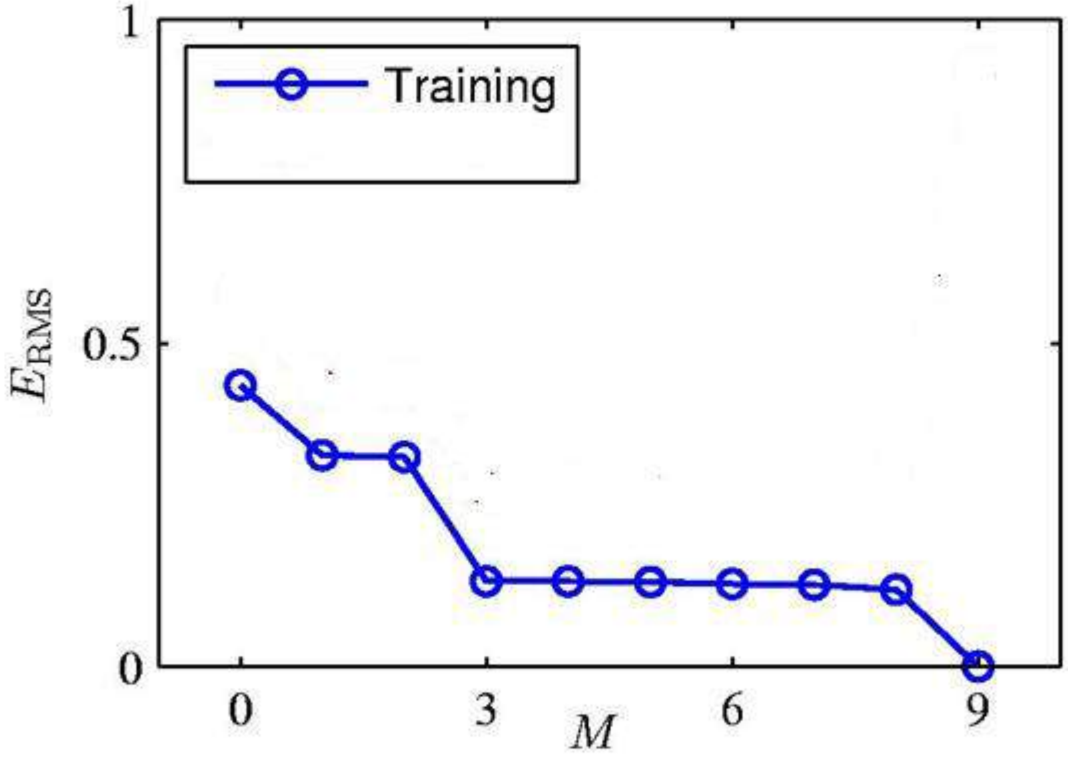
Model Selection - 3^{rd} Order Polynomial



Model Selection - 9th Order Polynomial

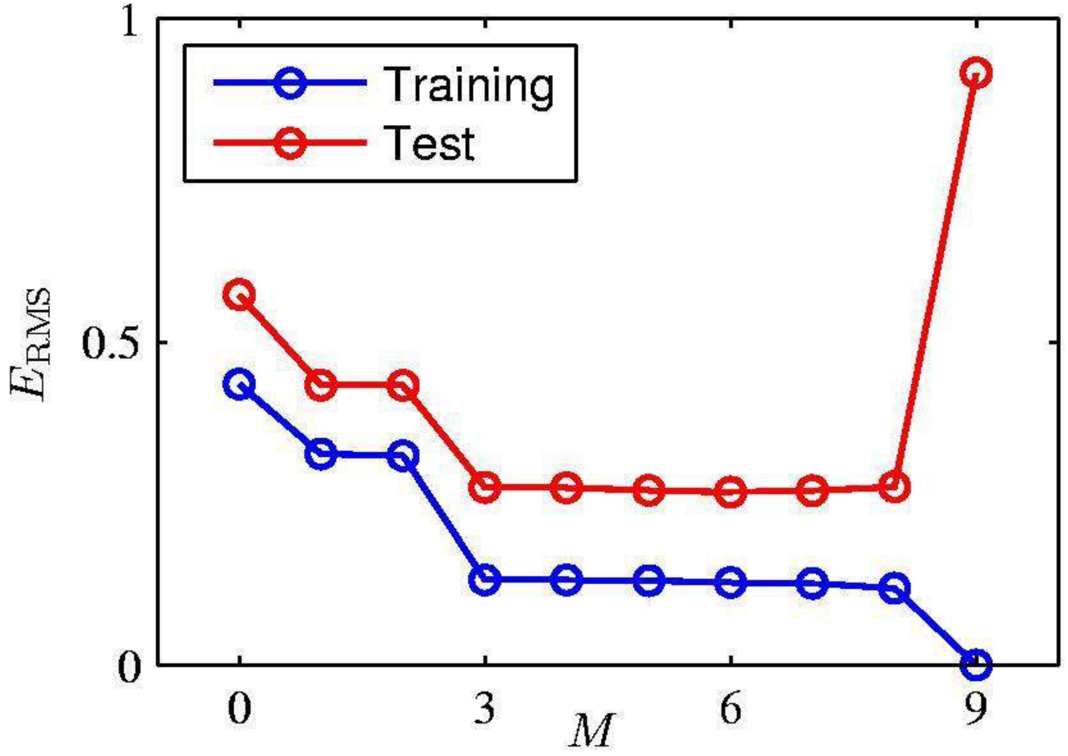


Model Selection



Root-Mean-Square (RMS) Error: $E_{\mathrm{RMS}} = \sqrt{2E(\mathbf{w}^\star)/N}$

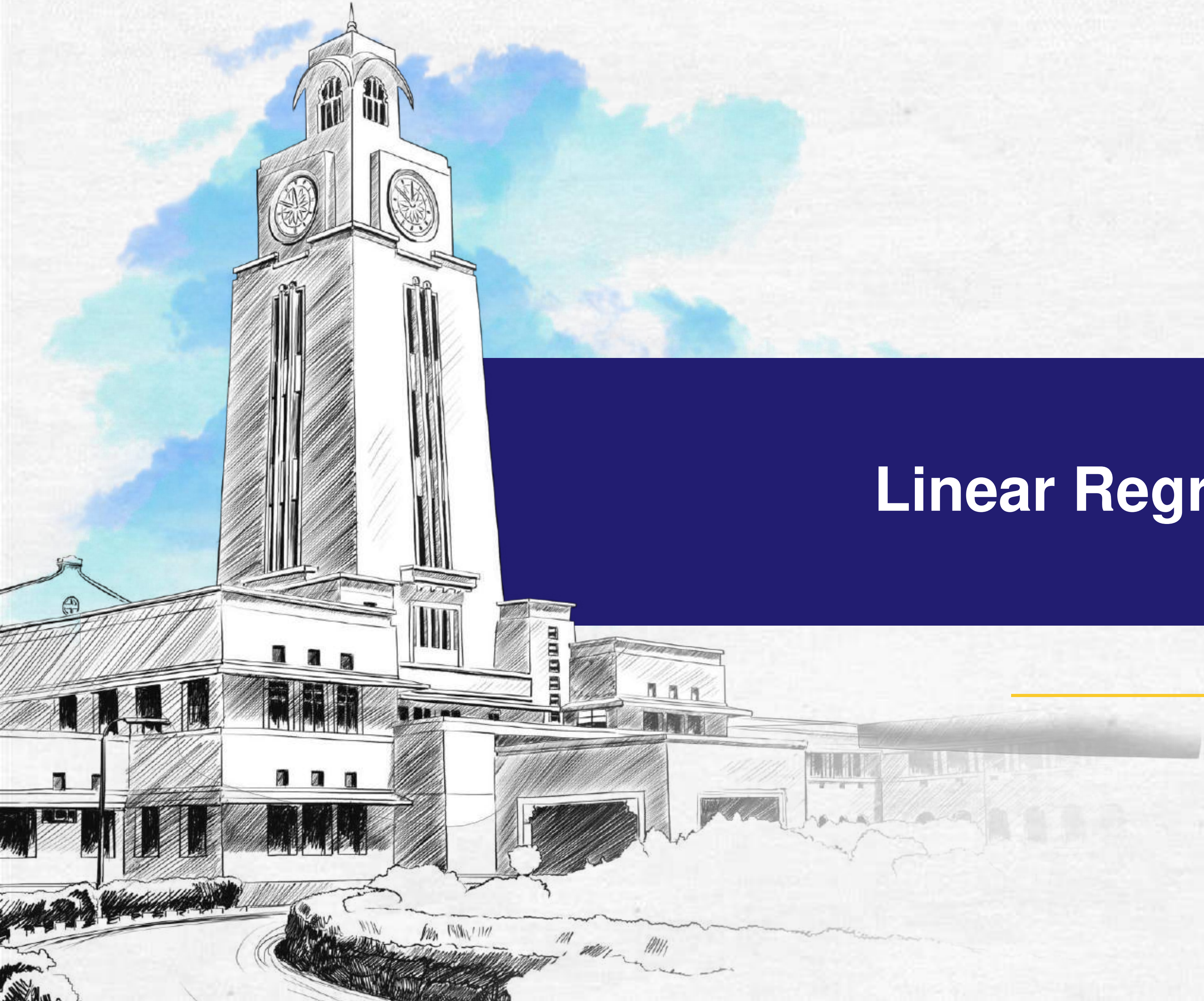
Model Selection



Root-Mean-Square (RMS) Error: $E_{\mathrm{RMS}} = \sqrt{2E(\mathbf{w}^\star)/N}$

Model Selection

Thank You!



BITS Pilani

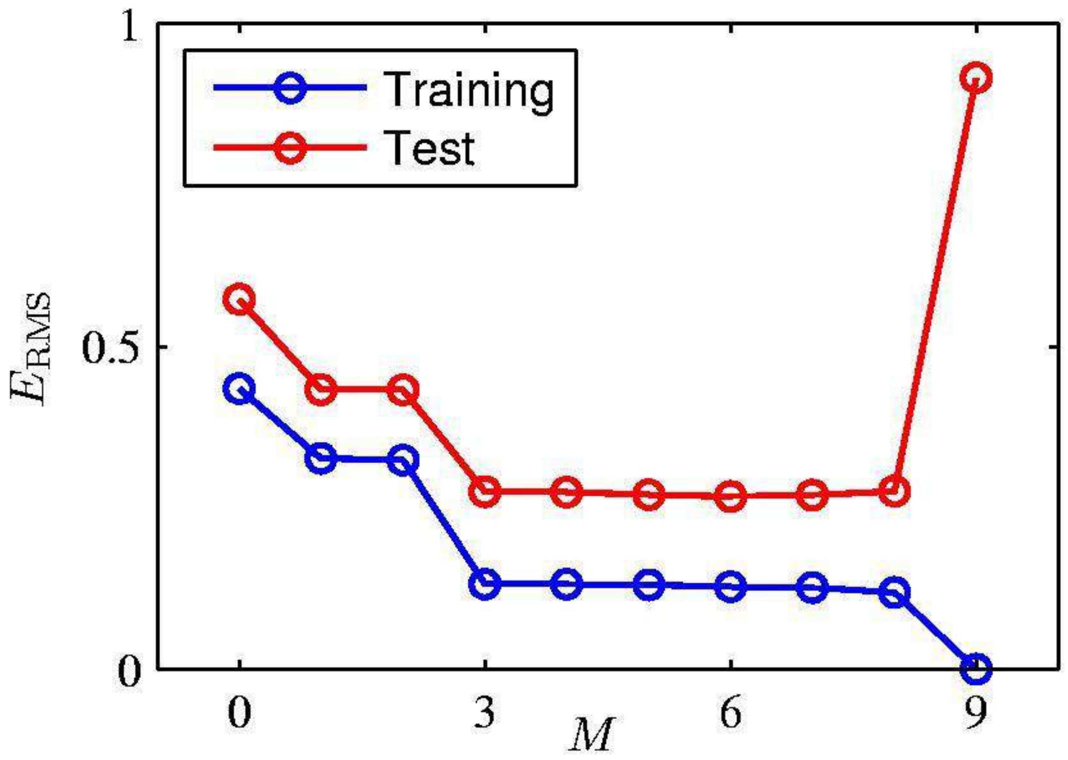
Pilani | Dubai | Goa | Hyderabad

Linear Regression - Overfitting

Prof.N.L.Bhanu Murthy

Overfitting

Overfitting



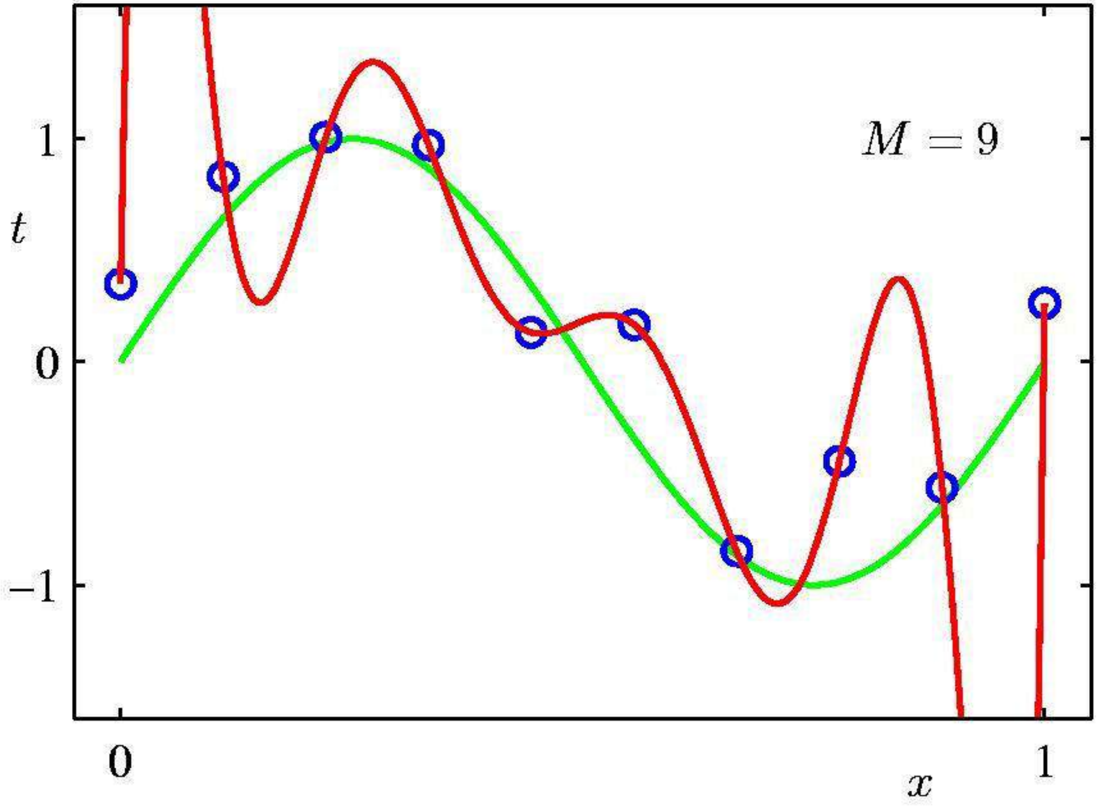
Overfitting

Polynomial Coefficients

	$M = 0$	$M = 1$	$M = 3$	$M = 9$
$w_0^\star$	0.19	0.82	0.31	0.35
$w_1^\star$		-1.27	7.99	232.37
$w_2^\star$			-25.43	-5321.83
$w_3^\star$			17.37	48568.31
$w_4^\star$				-231639.30
$w_5^\star$				640042.26
$w_6^\star$				-1061800.52
$w_7^\star$				1042400.18
$w_8^\star$				-557682.99
$w_9^\star$				125201.43

Overfitting

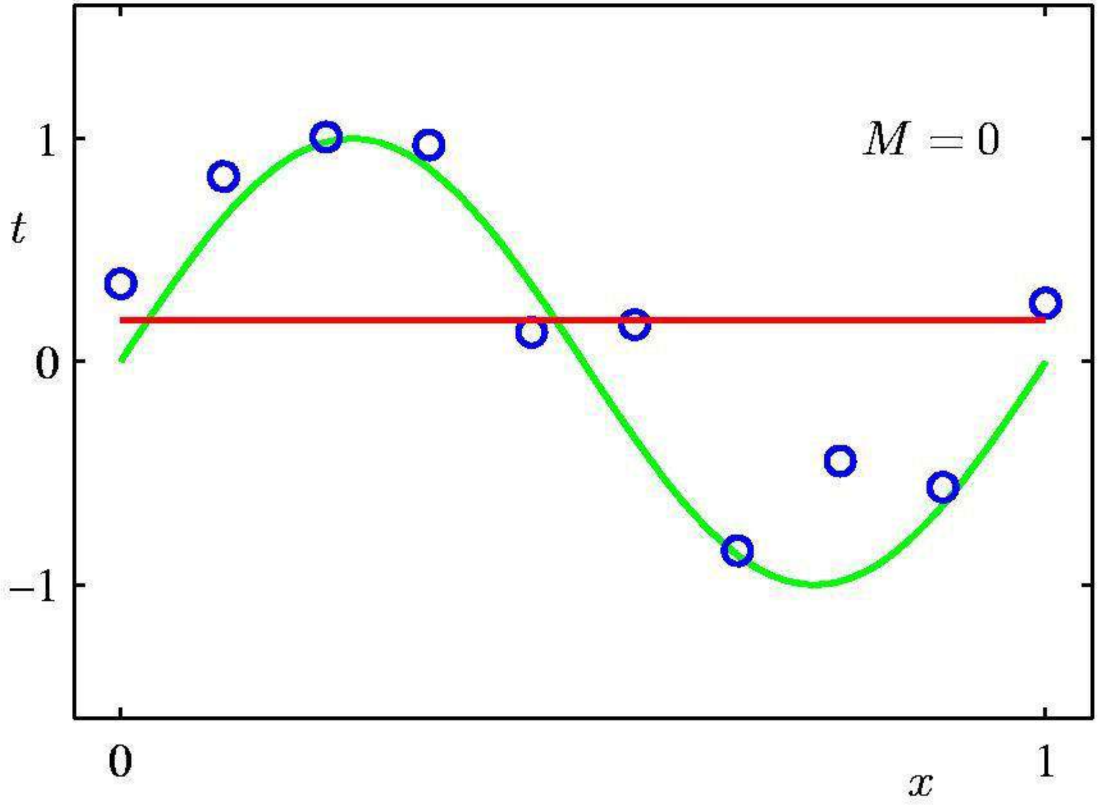
9^{th} Order Polynomial



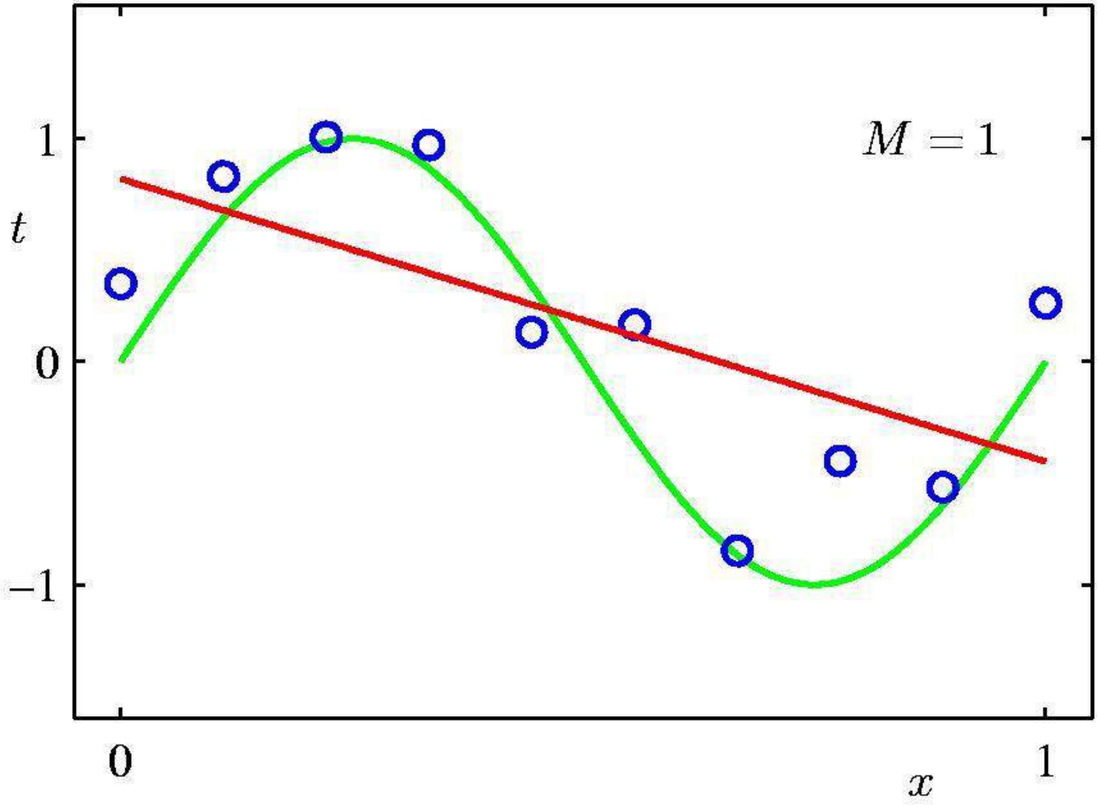
Overfitting

Overfitting

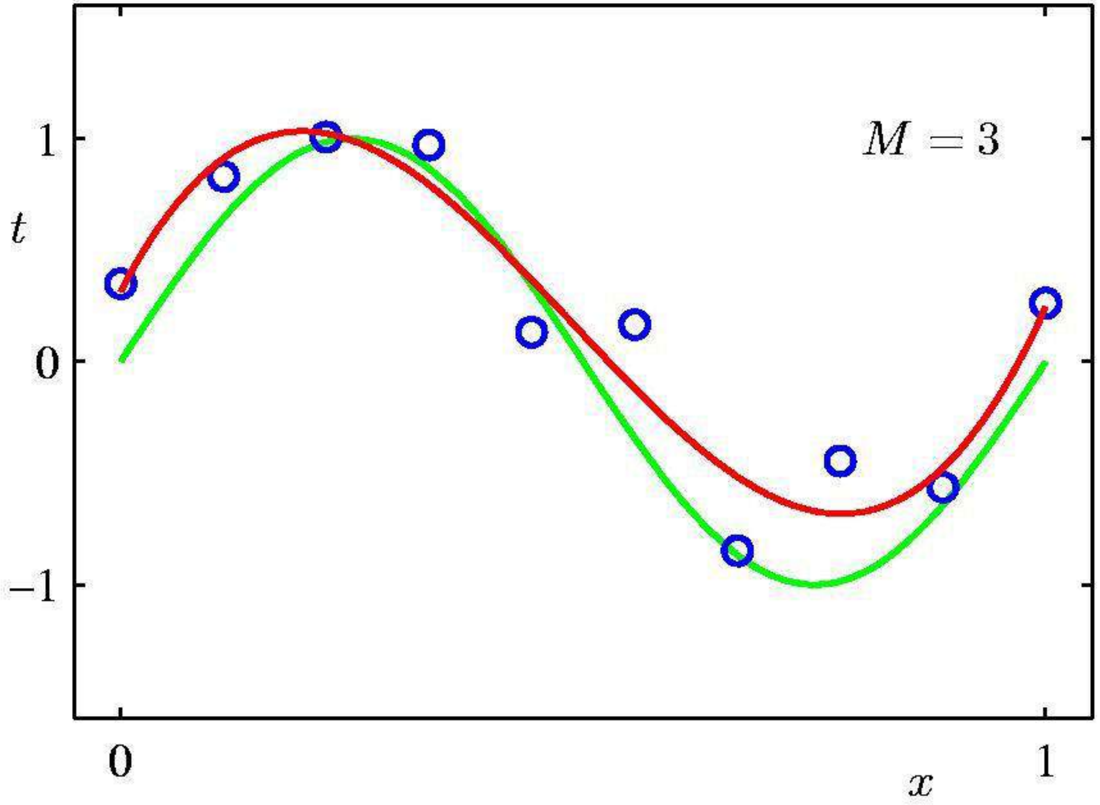
Model Selection - 0^{th} Order Polynomial



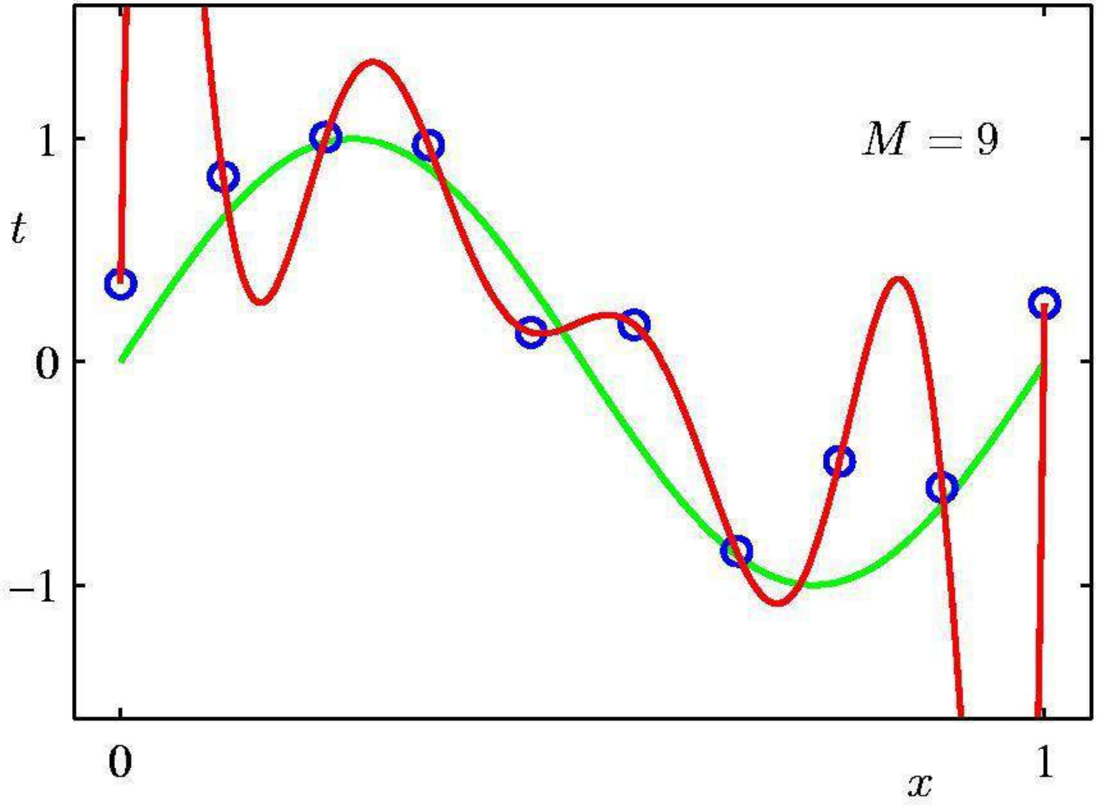
Model Selection - 1st Order Polynomial



Model Selection - 3rd Order Polynomial



Model Selection - 9th Order Polynomial



Thank You!

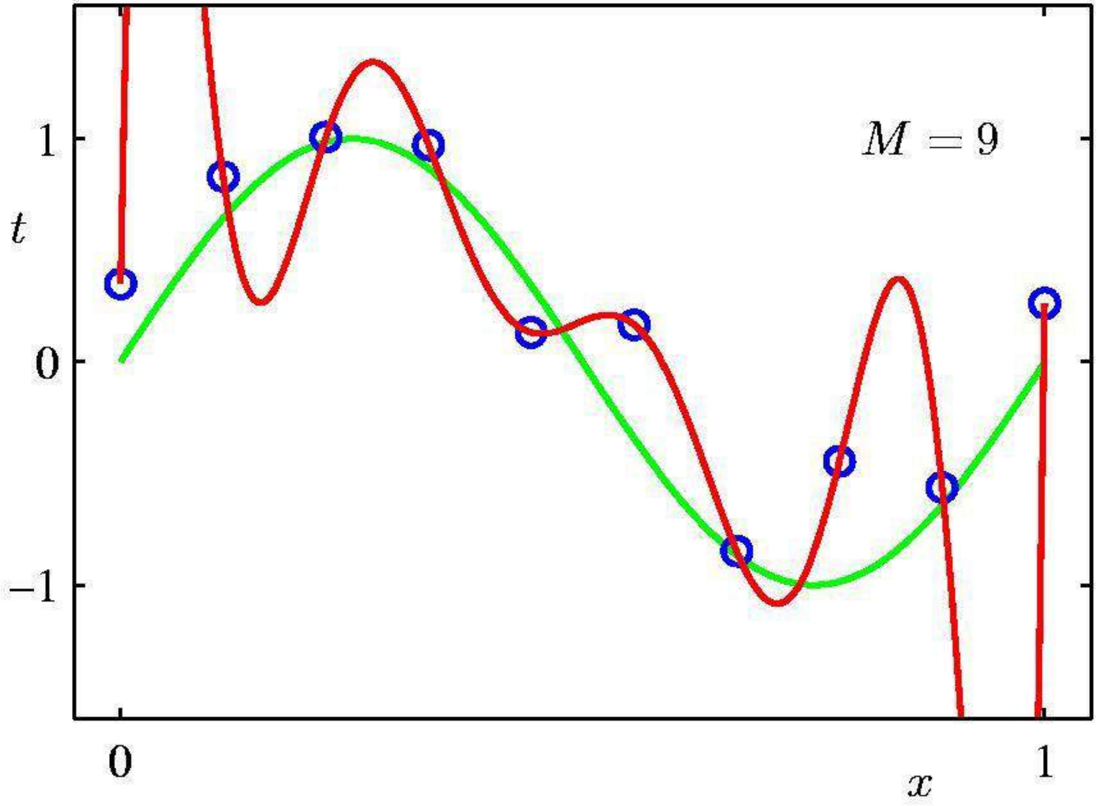
BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Linear Regression - Regularization

Prof.N.L.Bhanu Murthy

Model Selection - 9th Order Polynomial

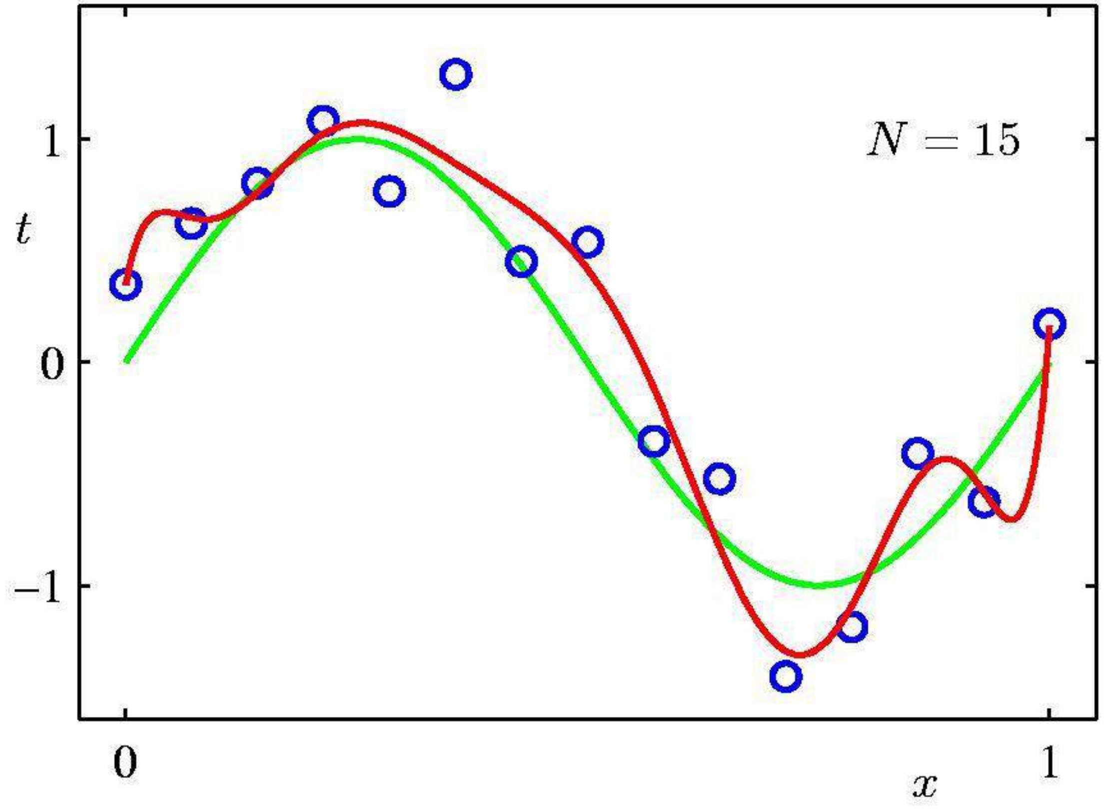
Data Set Size: N = 10



Overfitting

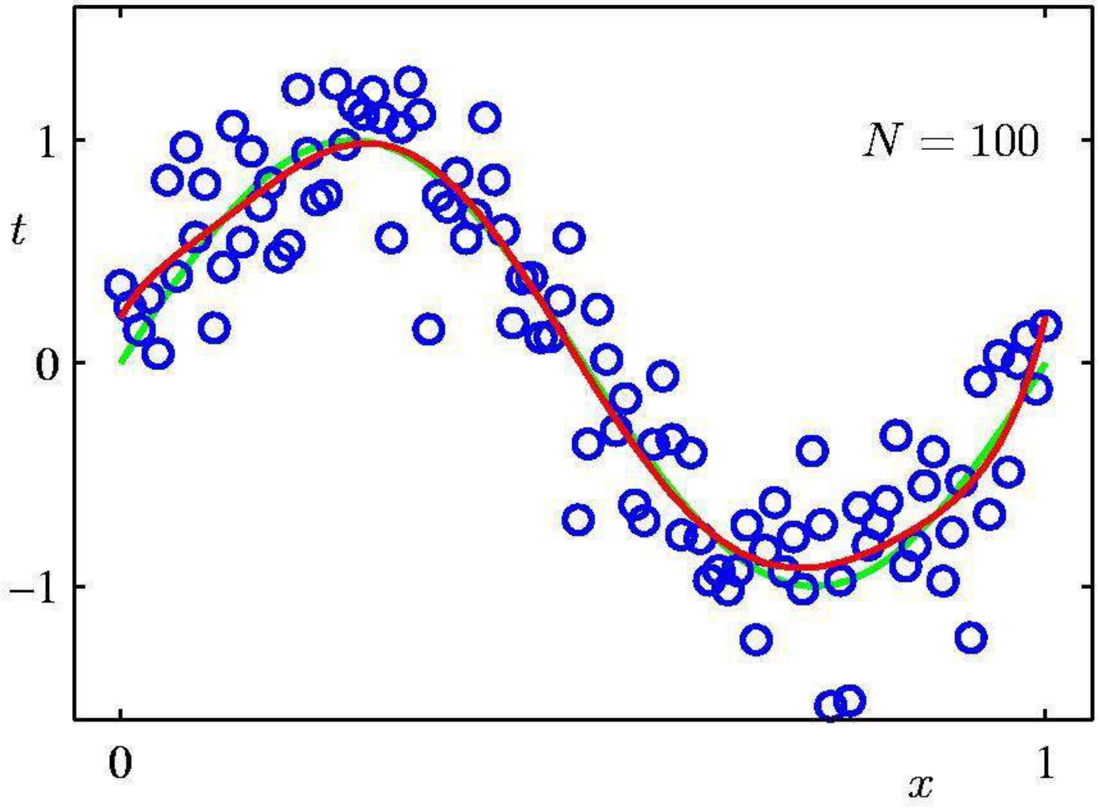
Data Set Size: $N = 15$

9^{th} Order Polynomial



Data Set Size: $N = 100$

9th Order Polynomial

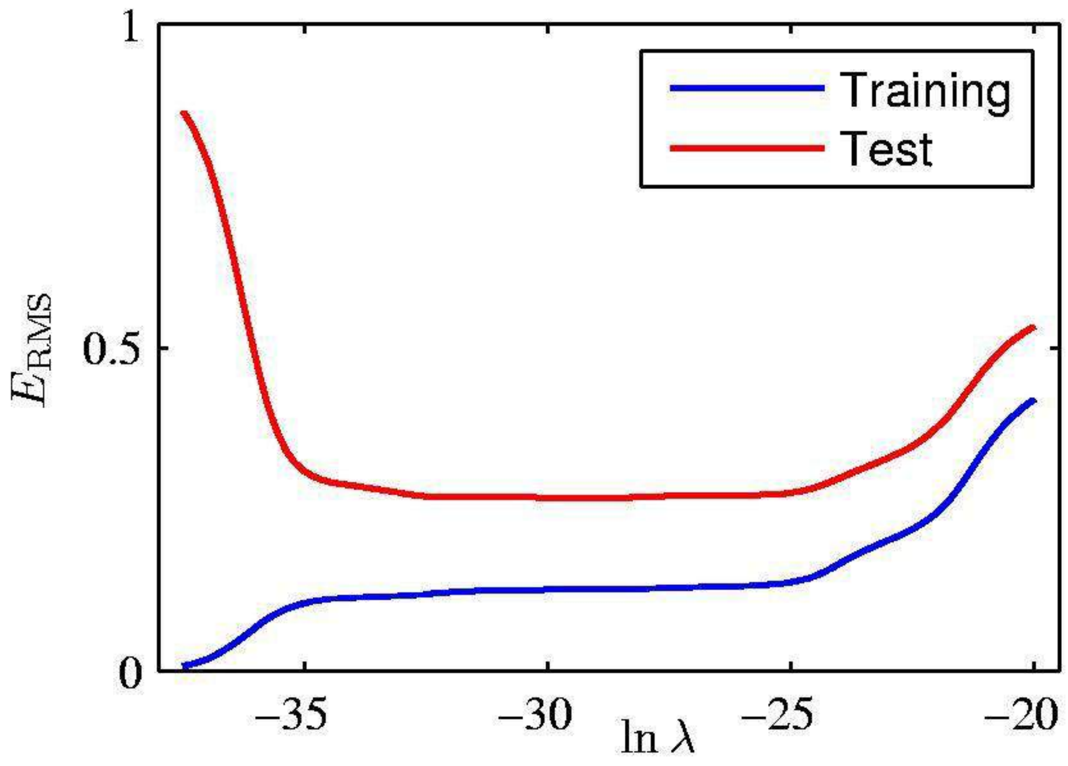


Overfitting

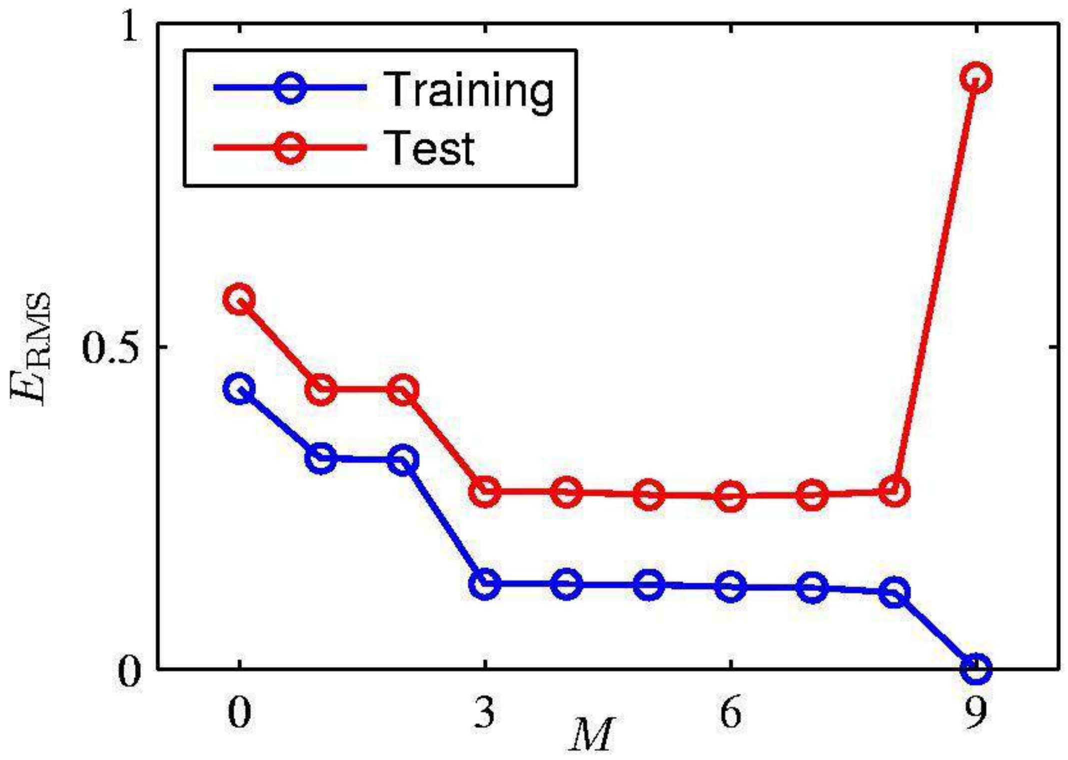
Polynomial Coefficients

	$M = 0$	$M = 1$	$M = 3$	$M = 9$
$w_0^\star$	0.19	0.82	0.31	0.35
$w_1^\star$		-1.27	7.99	232.37
$w_2^\star$			-25.43	-5321.83
$w_3^\star$			17.37	48568.31
$w_4^\star$				-231639.30
$w_5^\star$				640042.26
$w_6^\star$				-1061800.52
$w_7^\star$				1042400.18
$w_8^\star$				-557682.99
$w_9^\star$				125201.43

Regularization: E_{RMS} vs. $\ln \lambda$

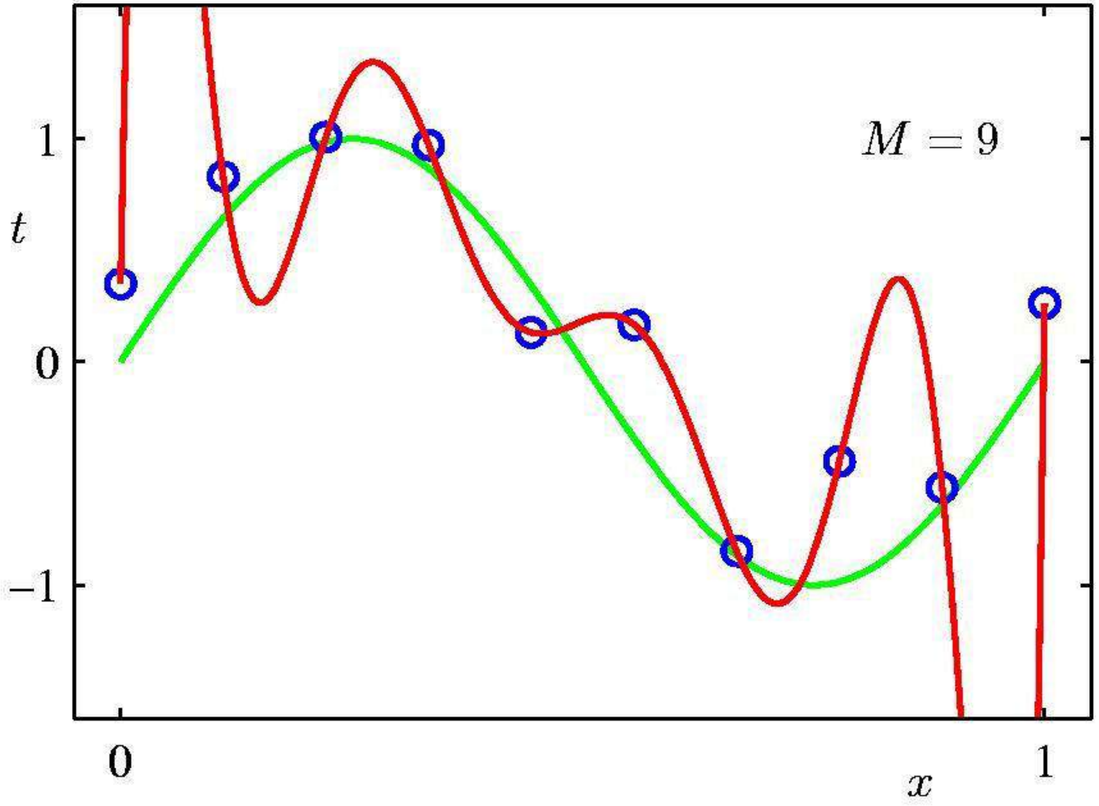


Overfitting



Overfitting

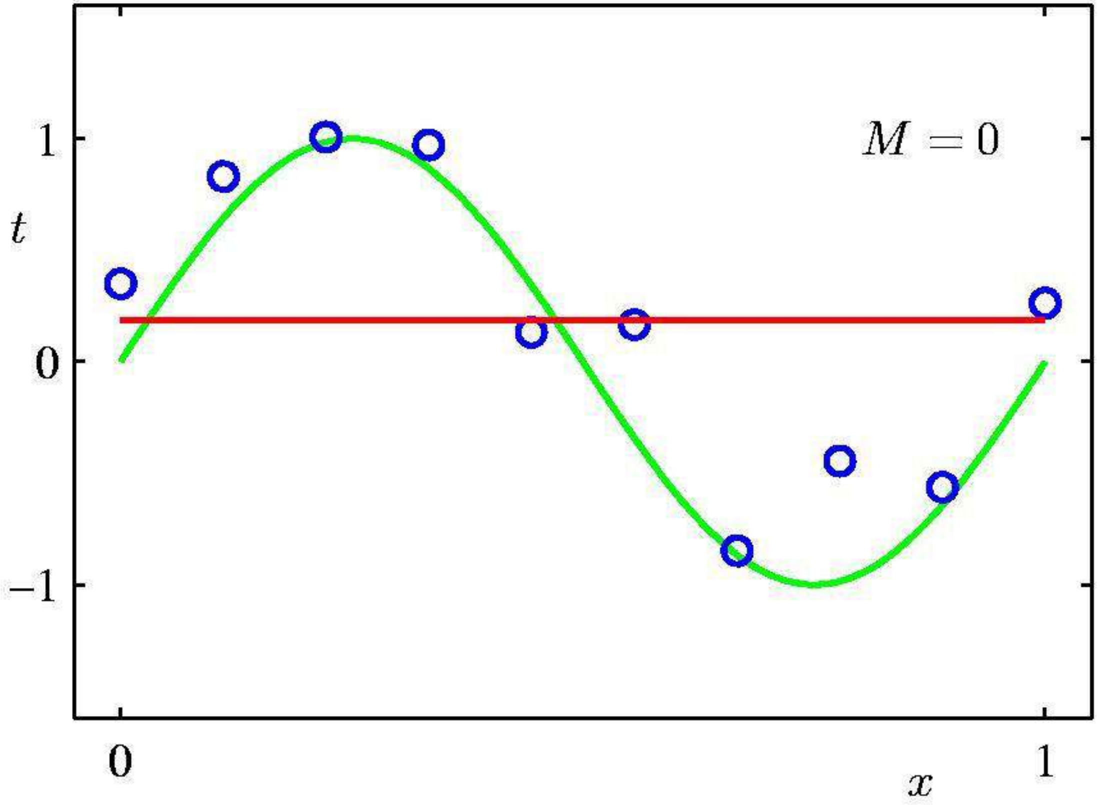
9th Order Polynomial



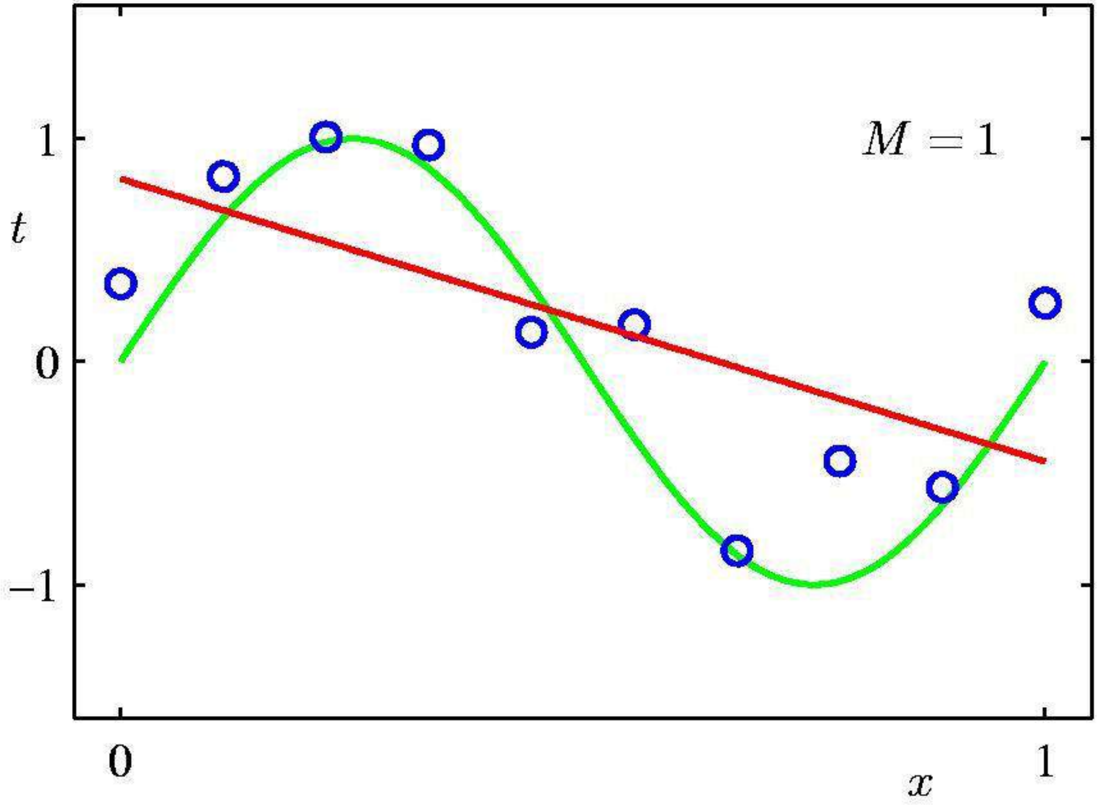
Overfitting

Overfitting

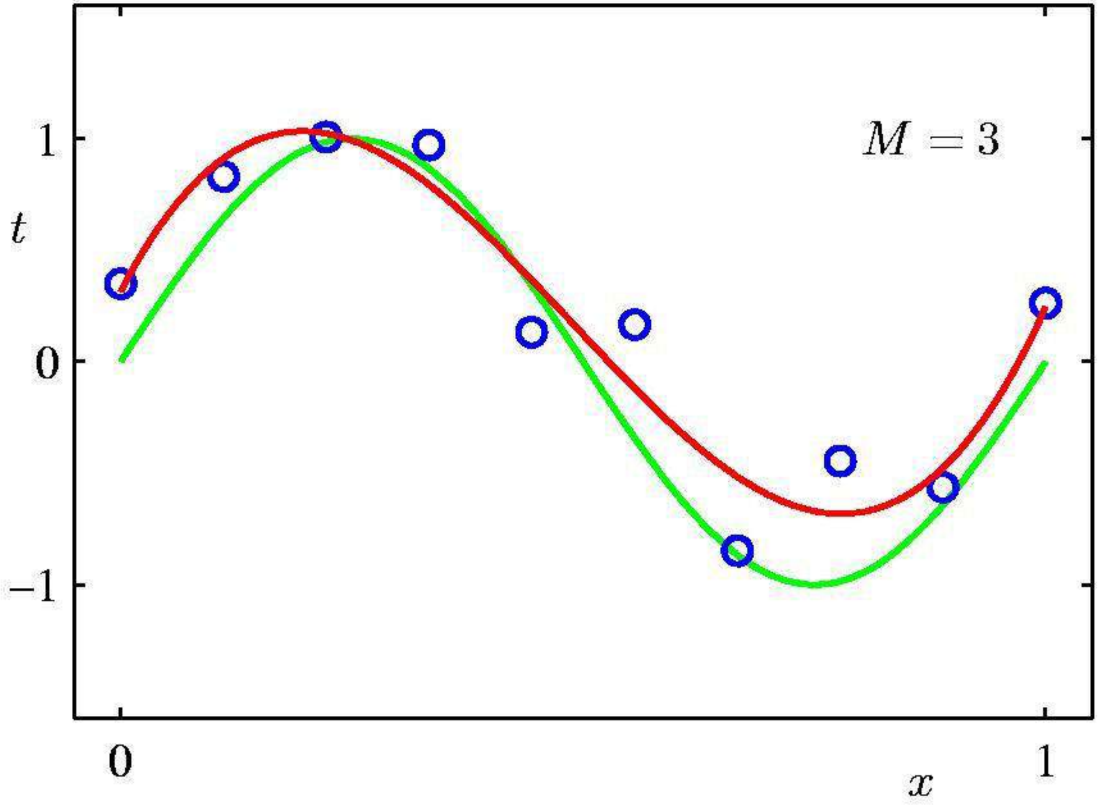
Model Selection - 0^{th} Order Polynomial



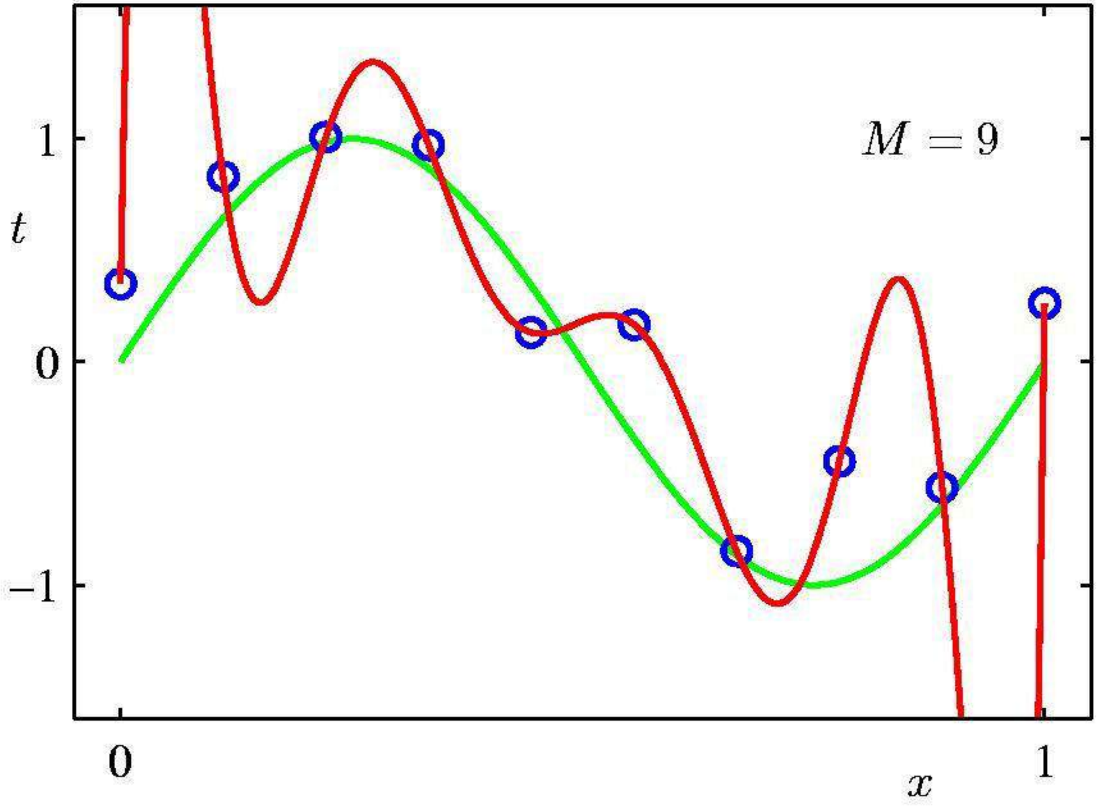
Model Selection - 1st Order Polynomial



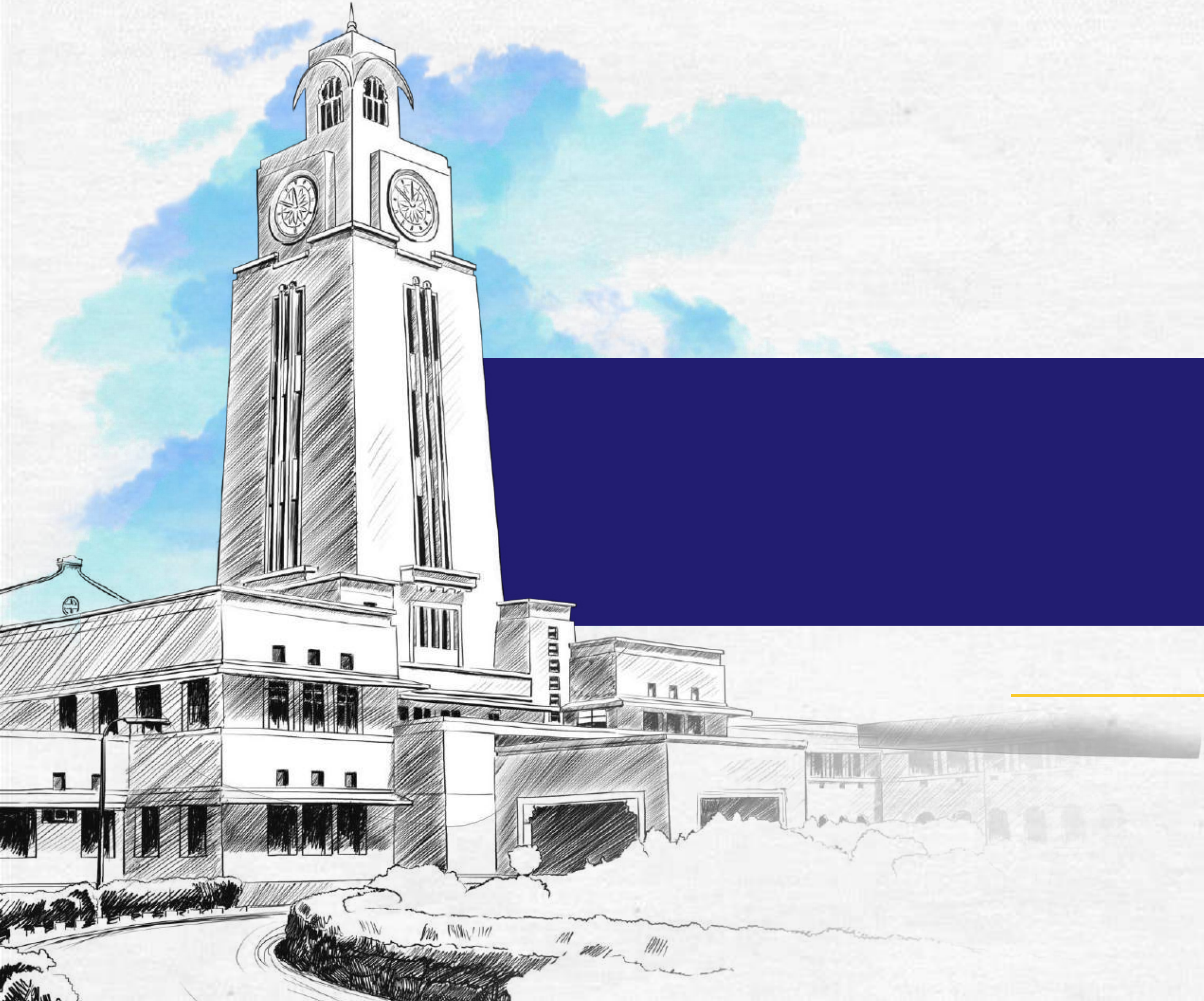
Model Selection - 3rd Order Polynomial



Model Selection - 9th Order Polynomial



Thank You!



BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Lasso Regression

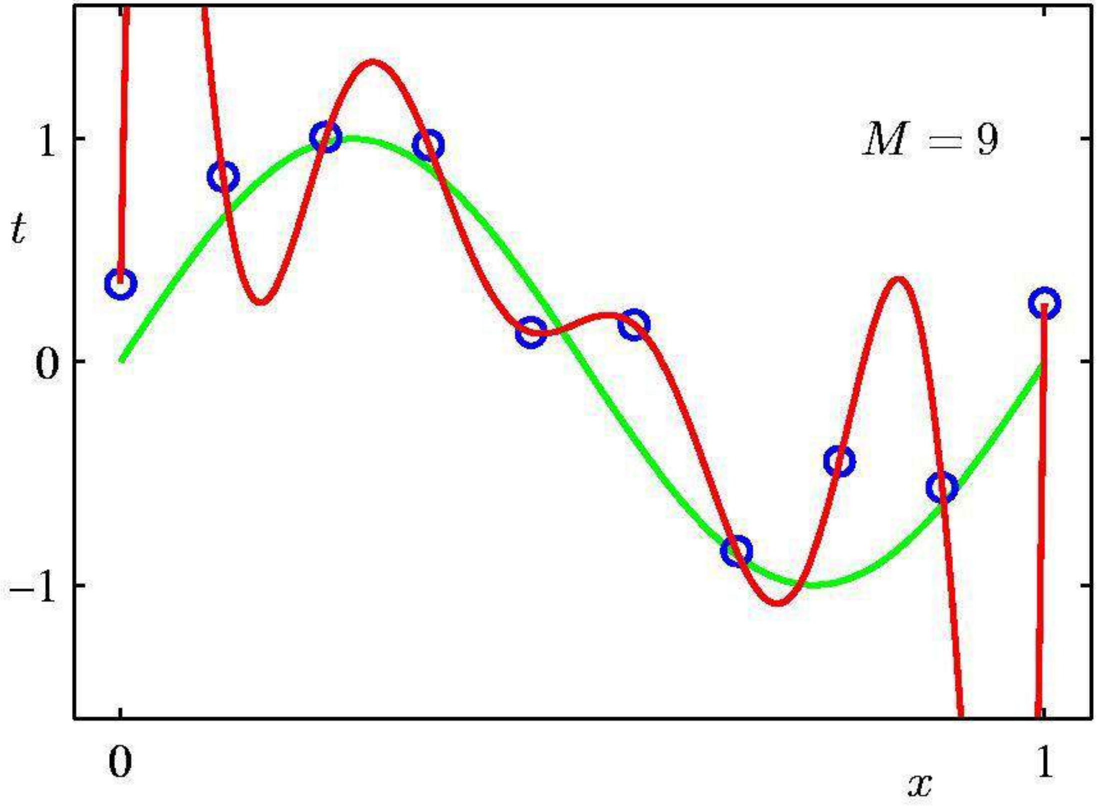
Prof.N.L.Bhanu Murthy

Regularization

Regularization

Regularization

Regularization



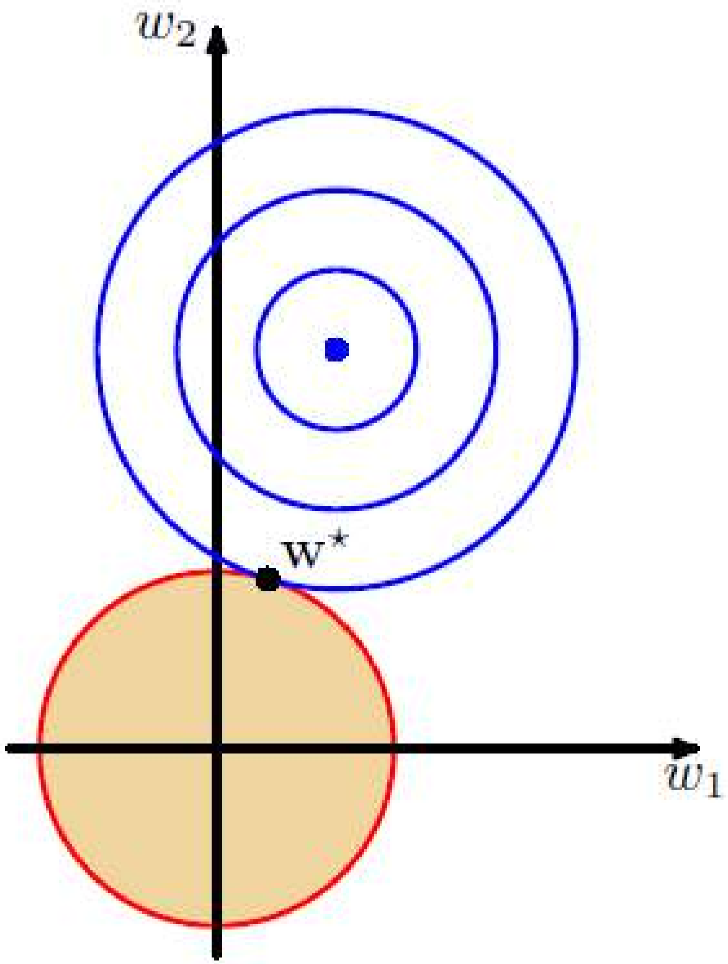
Regularization

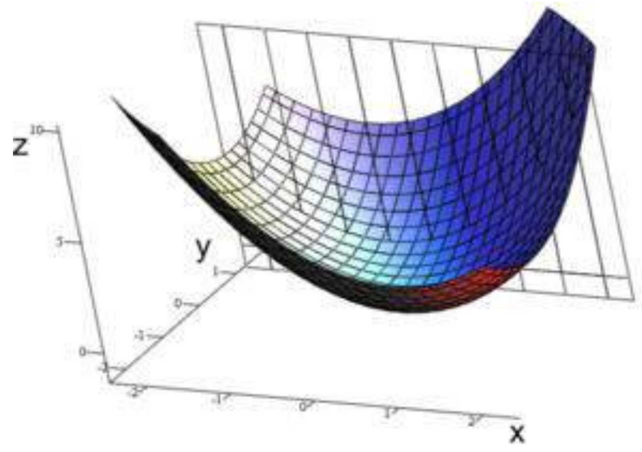
Polynomial Coefficients

	$M = 0$	$M = 1$	$M = 3$	$M = 9$
$w_0^\star$	0.19	0.82	0.31	0.35
$w_1^\star$		-1.27	7.99	232.37
$w_2^\star$			-25.43	-5321.83
$w_3^\star$			17.37	48568.31
$w_4^\star$				-231639.30
$w_5^\star$				640042.26
$w_6^\star$				-1061800.52
$w_7^\star$				1042400.18
$w_8^\star$				-557682.99
$w_9^\star$				125201.43

Regularization

Regularization

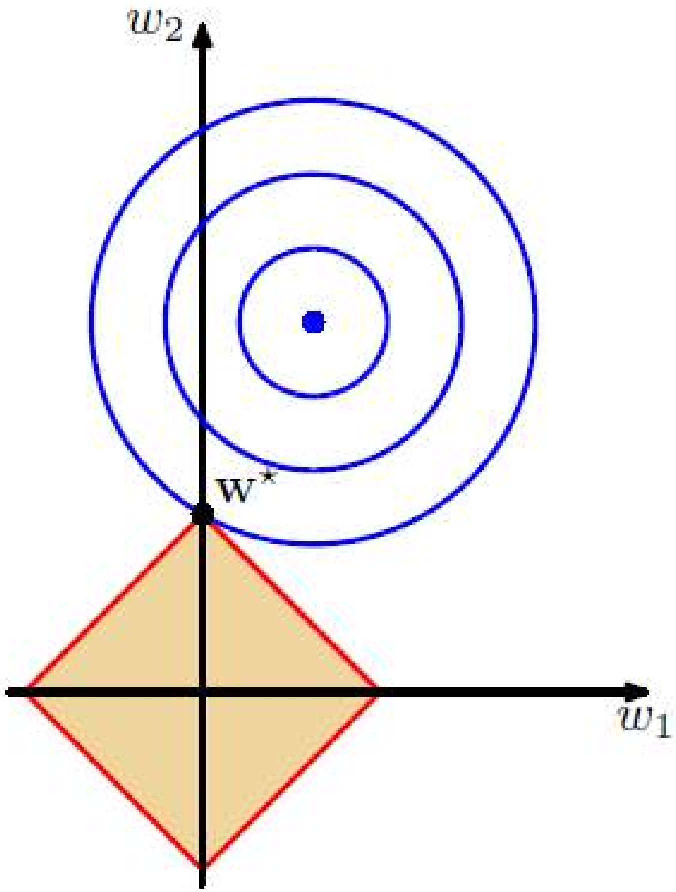


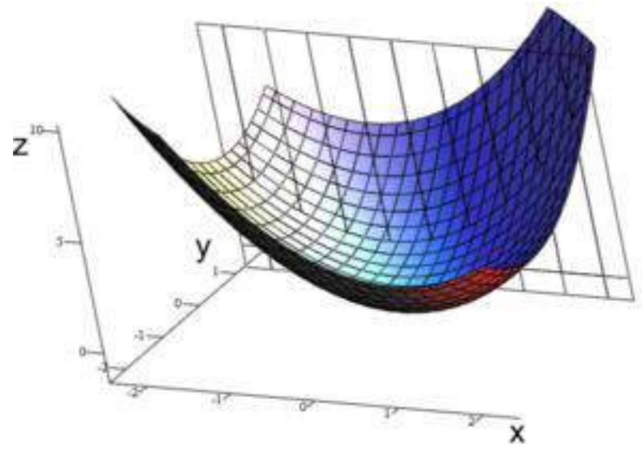


Regularization

Regularization

Lasso Regression





Regularization

Thank You!

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Linear Regression: Bias–Variance

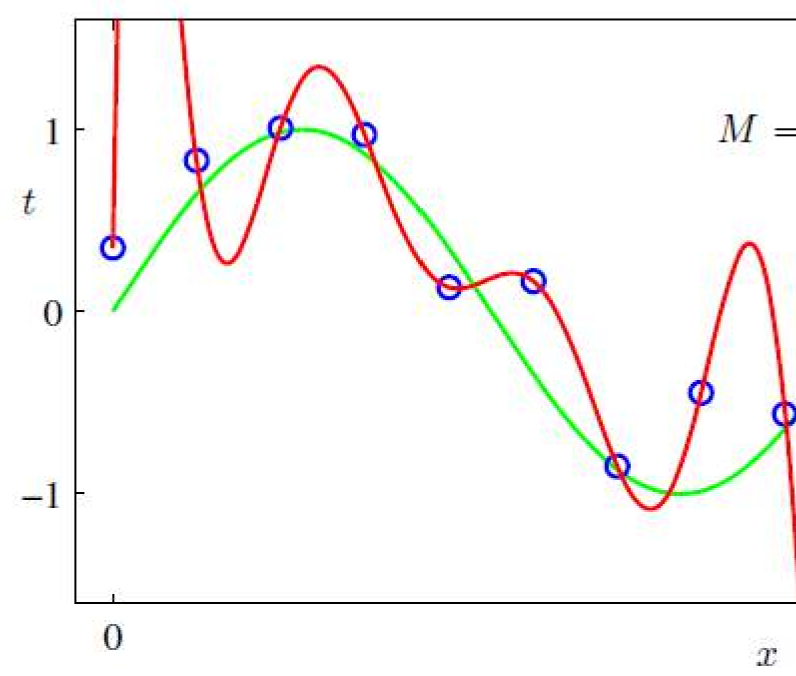
Prof.N.L.Bhanu Murthy

Bias-Variance

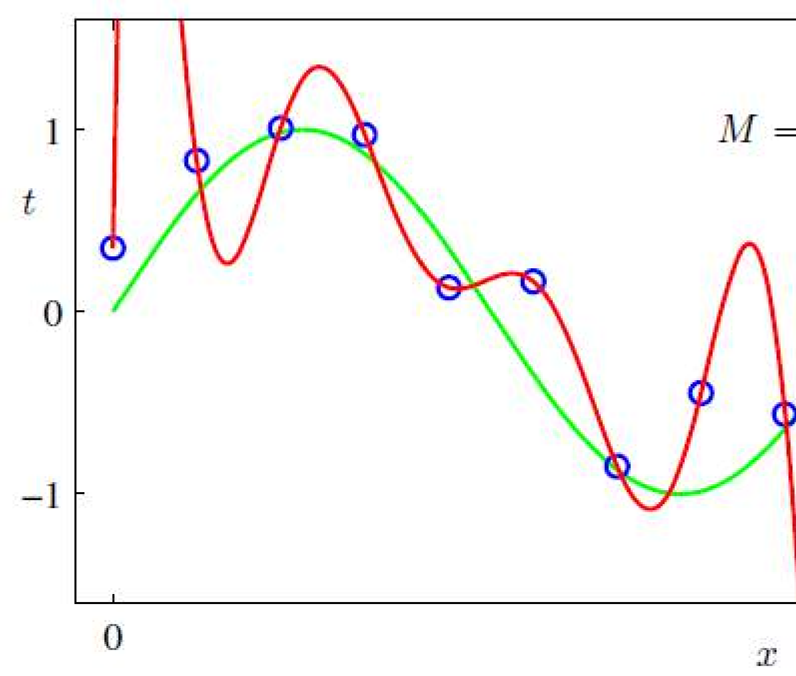
Bias-Variance

Bias-Variance

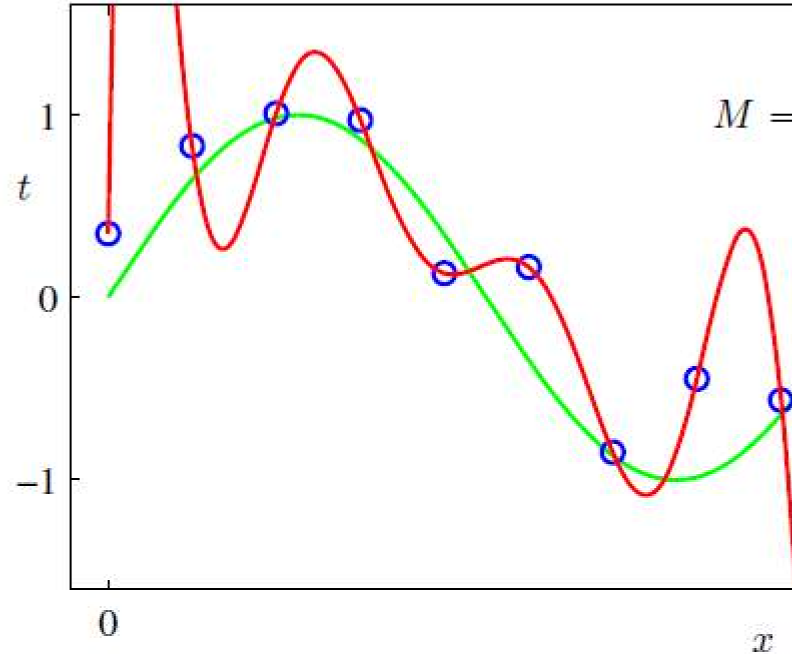
Bias-Variance



Bias-Variance



Bias-Variance



Bias-Variance

Bias-Variance

Bias-Variance

Bias-Variance

Bias-Variance

Bias-Variance

Bias-Variance

Bias-Variance

Bias-Variance

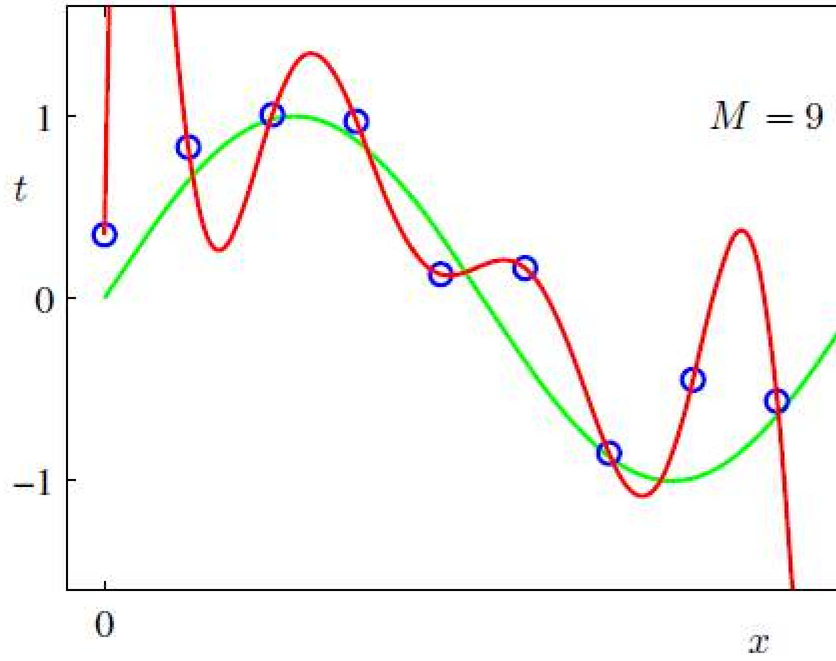
Bias-Variance

Thank You!

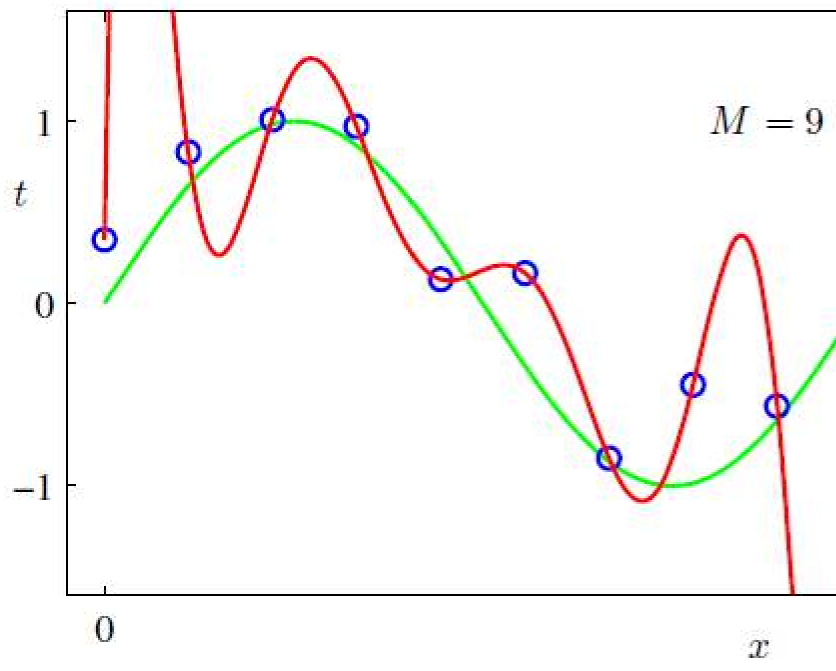
Linear Regression: Bias–Variance

Prof.N.L.Bhanu Murthy

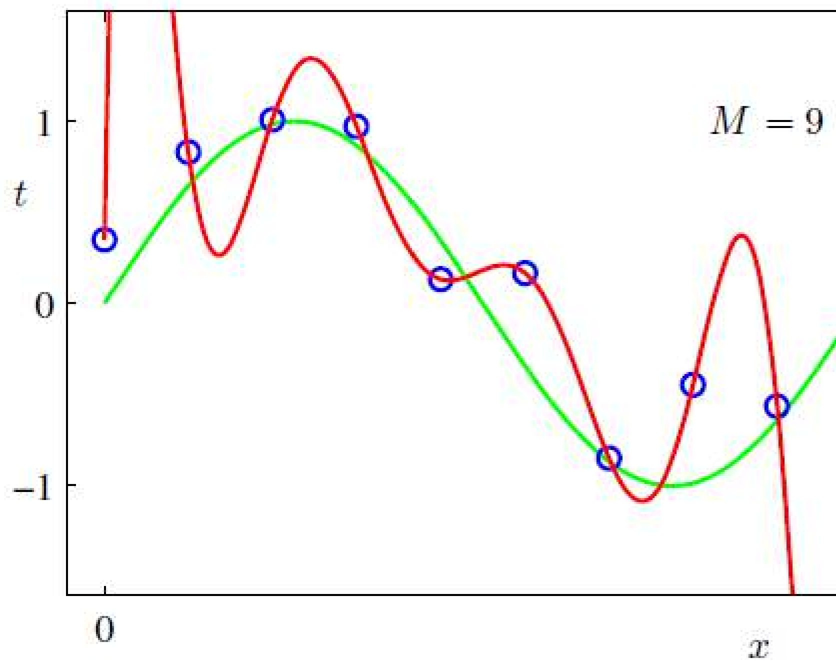
Bias-Variance



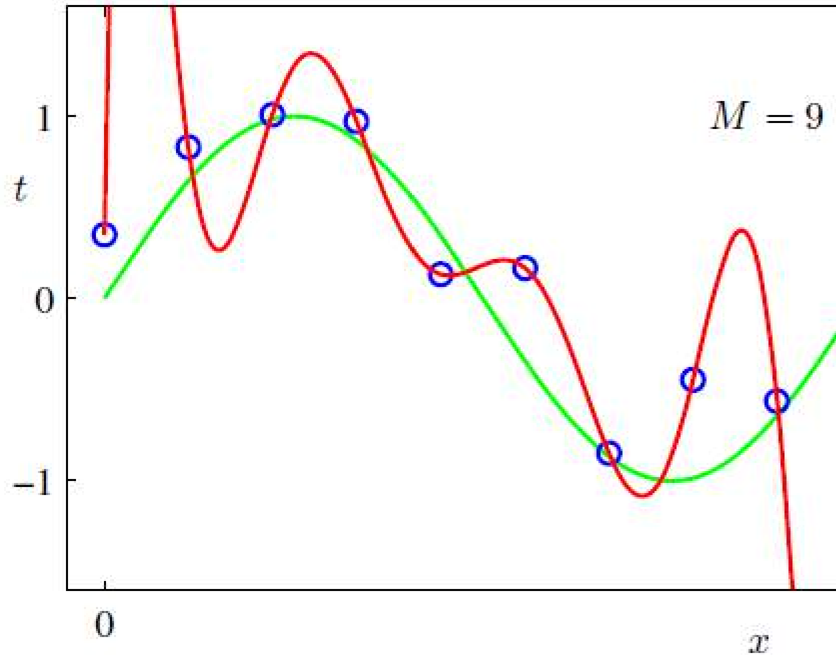
Bias-Variance



Bias-Variance



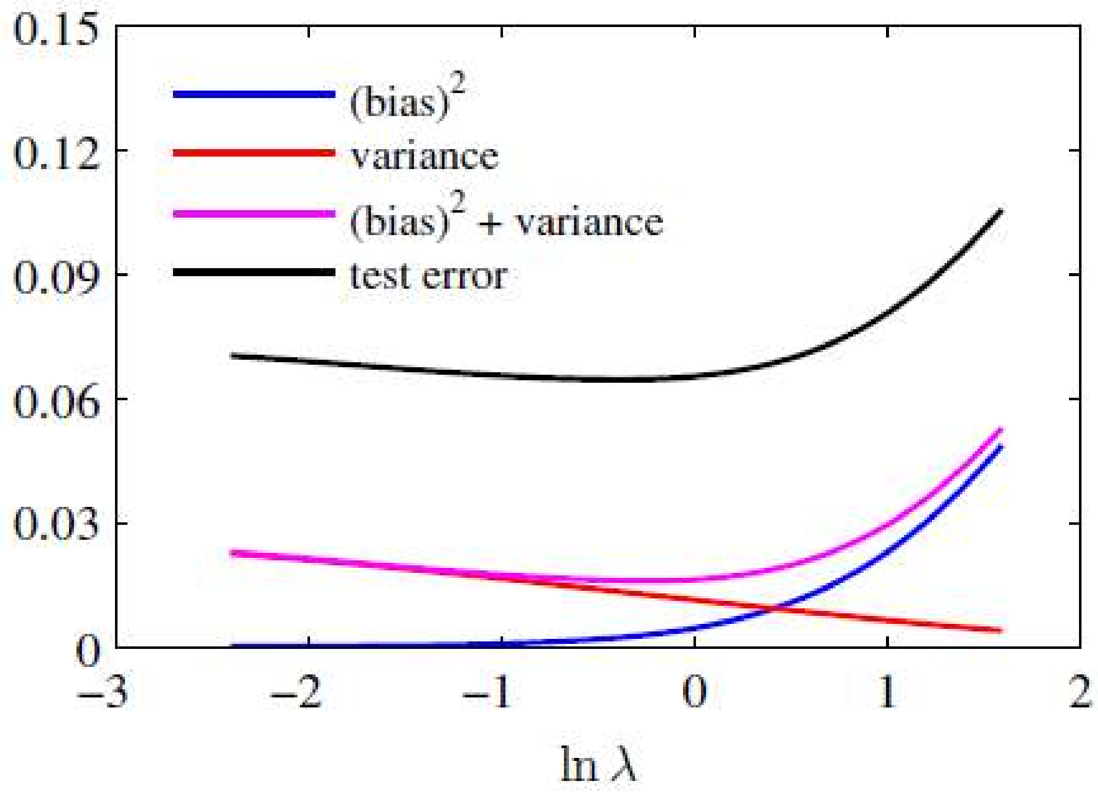
Bias-Variance



Bias-Variance

Bias-Variance

Bias-Variance



Bias-Variance

Bias-Variance

Bias-Variance

Bias-Variance

Bias-Variance

Bias-Variance

Bias-Variance

Bias-Variance

Bias-Variance

Bias-Variance

Bias-Variance

Bias-Variance

Bias-Variance

Thank You!