Foundation of Machine Learning

CSE4032

Lecture 01: Introduction

Dr. Kundan Kumar
Associate Professor
Department of ECE

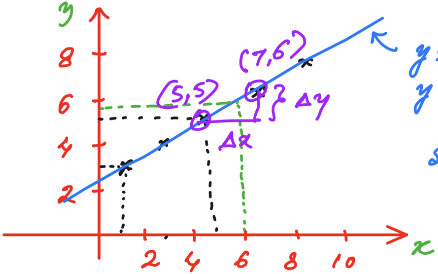
Faculty of Engineering (ITER)
S'O'A Deemed to be University, Bhubaneswar, India-751030

Better to start with examples

apple	cost(Rs)
x	y
1	3
3	4
5	5
7	6
9	7
6	? 5.5

$$\boxed{y = 0.5x + 2.5}$$

50 apple



Regression

$y = 0.5x + 2.5$

$y = mx + c$

$y = a_1 x + a_0$

slope

intercept

$$a_1 = m = \frac{\Delta y}{\Delta x}$$

$$= \frac{y_2 - y_1}{x_2 - x_1}$$

$$= \frac{1}{2} = 0.5$$

$a_0 = 2.5$

Better to start with examples

"x" Hours of Sunshine	"y" Ice Creams Sold
x_1 2	4
x_2 3	5
x_3 5	7
x_4 7	10
x_5 9	15

$$y = a_1 x + a_0$$

11 hr.

$$\hat{y}_5 = a_1 \times 9 + a_0$$

$$y_5 = 15$$

$$E = E_1 + E_2 + E_3 + E_4 + E_5$$

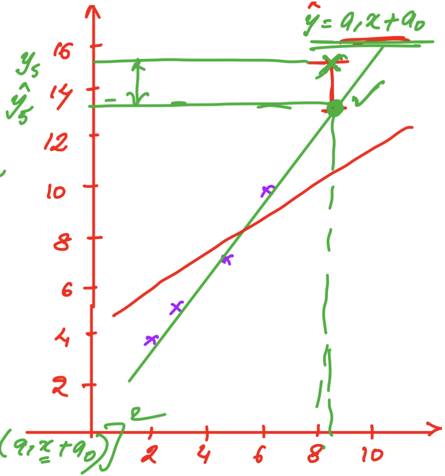
$$E = \sum_{i=1}^{n} |y - \hat{y}|$$

SSE

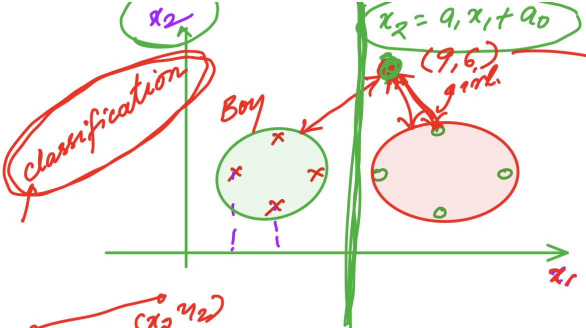
$$= \sum_{i=1}^{n} (y - \hat{y})^2$$

$$= E = \sum_{i=1}^{n} [y - (a_1 x + a_0)]^2$$

$$\frac{\partial E}{\partial a_1} = 0 \quad \frac{\partial E}{\partial a_0} = 0$$



Better to start with examples



x_1	x_2	y
2	2	w_1
3	1	w_1
3	3	w_1
4	2	w_1
6	2	w_2
8	1	w_2
8	3	w_2
10	2	w_2

(x_1, y_1) $(x_2 y_2)$

Euclidean Distance

$$= \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2}$$

Outline

1. Course Details
2. Statistical Learning
3. Example of SL
4. References

Google Classrooms

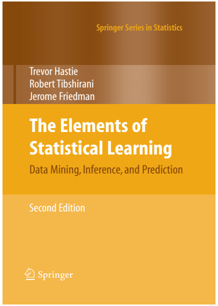
- All the communication will be through Google Classroom:
  - Course materials
  - Assignments
  - Announcements and Notices
- Join the course at `https://classroom.google.com/`
- Joining classroom request will be sent to the email ids or student may joint through the following course code.
- Class code for CSE, Section-C:

  ahxoqy5

- Class code for CSE, Section-D:

  t76ibfr

Text Books



Springer Series in Statistics

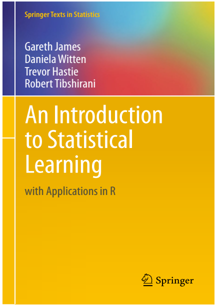
Trevor Hastie
Robert Tibshirani
Jerome Friedman

The Elements of Statistical Learning

Data Mining, Inference, and Prediction

Second Edition

Springer



Springer Texts in Statistics

Gareth James
Daniela Witten
Trevor Hastie
Robert Tibshirani

An Introduction to Statistical Learning

with Applications in R

Springer

Text Books:

- The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, Hastie, Tibshirani, and Friedman, Springer. ▸ download
- An Introduction to Statistical Learning with Application in R, Second Edition, James, Witten, Hastie, and Tibshirani, Springer. ▸ download

Credits:

- 4 credits course,
  - □ 3 classes/week (1hr/class),
  - □ 1 problem solving (2hr session/week).

Grading pattern

- Grading pattern: 1

Attendance	:	5 Marks
2 Quizzes	:	10 Marks
Assignments	:	10 Marks
Mid-term examination	:	15 Marks
Total Internal	:	40 Marks

In lab exam	:	15 Marks
Theory exam	:	45 Marks
Total External	:	60 Marks

Regarding Attendance and Home Assignments

- Attendance will be taken by calling students name or last three digit of the registration number.
- Alternatively, attendance will be taken through the Google Form.

In this course we are going to cover

- Overview of Supervised Learning
- Linear Models for Regression
- Linear Models of Classification
- Basic Expansion and Regularisation
- Kernel Smoothing Methods
- Model Assessment and Selection
- Model Inference and Averaging
- Additive Modes
- Trees and Related Methods
- Neural Networks

Introduction to statistical learning

- Introduced in the late 1960s.
- It was simply a problem of function estimation form a given collection of data.
- In the middle of 1990s, a new types of learning algorithm (e.g. SVM) based on developed theory were developed.
- This made statistical learning theory not only a tool for the theoretical analysis but also a tool for creating practical algorithm for estimating multidimensional functions.

Statistical Learning

- Statistical learning deals with learning from data.
- A typical scenario: we have outcome measurement
  - □ quantitative, e.g., stock price
  - □ categorical, e.g. heart attack/no heart attack
- Based on set of features, e.g., historical price, diet plan, clinical measurement,a good learner accurately predicts the outcomes. This learning process is called supervised learning.
- Training set - used to observe the outcome and feature measurements for a set of objects.
  - □ Using this set, we build a prediction model, or a statistical learner, which enables us to predict the outcome for a set of new unseen objects.

Good learner

- A good learner is one that accurately predicts such an outcome.
- All statistical learning problems may be constructed by minimizing the expected loss.
  - Mathematically, problem of learning is that of choosing from the given set of functions, the one that predicts the supervised learning's response in the best way.
  - In order to choose the best available response, a risk functional is minimized in a situation, where the joint distribution of the predictions and response is unknown and the only available information is obtained from the training data.

Statistical Learning vs. Machine Learning

- Machine learning arose as a subfield of Artificial Intelligence.
- Statistical learning arose as a subfield of Statistics.
- There is much overlap — both fields focus on supervised and unsupervised problems:
  - □ Machine learning has a greater emphasis on large scale applications and prediction accuracy.
  - □ Statistical learning emphasizes models and their interpretability, and precision and uncertainty.
- But the distinction has become more and more blurred, and there is a great deal of "cross-fertilization".
- Machine learning has the upper hand in Marketing!

What is Machine Learning?

- Machine Learning is concerned with the development of algorithms and techniques that allow computers to learn.
- Learning in this context is the process of gaining understanding by constructing models of observed data with the intention to use them for prediction.
- Related fields
  - Artificial Intelligence: smart algorithms
  - Statistics: inference from a sample
  - Data Mining: searching through large volumes of data
  - Pattern classification: feature engineering and standard algorithms.

Why ‘Learn‘?

- There is no need to “learn” to calculate payroll.
- Learning is used when:
  - Human expertise does not exist (navigating on Mars),
  - Humans are unable to explain their expertise (speech recognition).
  - Solution changes in time (routing on a computer network)
  - Solution needs to be adapted to particular cases (user biometrics).
- Example: It is easier to write a program that learns to play checkers or chess well by self-play rather than converting the expertise of a master player to a program.

Examples of statistical learning

Examples of learning problems

- Predict whether a patient, hospitalized due to a heart attack, will have a second heart attack. The prediction is to be based on demographic, diet and clinical measurements for that patient.
- Predict the price of a stock in 6 months from now, on the basis of company performance measures and economic data.
- Identify the numbers in a handwritten ZIP code, from a digitized image.
- Estimate the amount of glucose in the blood of a diabetic person, from the infrared absorption spectrum of that person's blood.
- Identify the risk factors for prostate cancer, based on clinical and demographic variables.

Example: Email Spam Classification

- This is a supervised learning problem, with the outcome the class variable email/spam. It is also called a classification problem.
- The data for this example consists of information from 4601 email messages, in a study to try to predict whether the email was junk email, or “spam.”
- The objective was to design an automatic spam detector that could filter out spam before clogging the users’ mailboxes.
- For all 4601 email messages, the true outcome (email type) email or spam is available, along with the relative frequencies of 57 of the most commonly occurring words and punctuation marks in the email message.

Example: Email Spam Classification

	george	you	your	hp	free	hpl	!	our	re	edu	remove
spam	0.00	2.26	1.38	0.02	0.52	0.01	0.51	0.51	0.13	0.01	0.28
email	1.27	1.27	0.44	0.90	0.07	0.43	0.11	0.18	0.42	0.29	0.01

- Our learning method has to decide which features to use and how: for example, we might use a rule such as

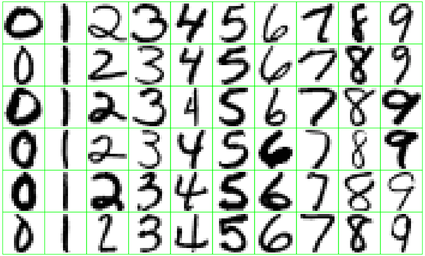
$$\begin{aligned}\text{if } (\%\text{ george } < 0.6)\ \ \&\ \ (\ \%\text{you } > 1.5)\ \text{then spam}\\ \text{else email.}\end{aligned}$$

- Another form of a rule might be:

$$\begin{aligned}\text{if } (0.2\cdot\%\text{ you } - 0.3\cdot\ \%\text{george }) > 0\ \text{then spam}\\ \text{else email.}\end{aligned}$$

Example: Handwritten Digit Recognition

- The data from this example come from the handwritten ZIP codes on envelopes from U.S. postal mail.
- The images are 16×16 eight-bit grayscale maps, with each pixel ranging in intensity from 0 to 255.
- The task is to predict, from the 16 × 16 matrix of pixel intensities, the identity of each image $(0, 1, \ldots, 9)$ quickly and accurately.



Example: Pixel Classification

- Classify the pixels in a LANDSAT image, by usage.

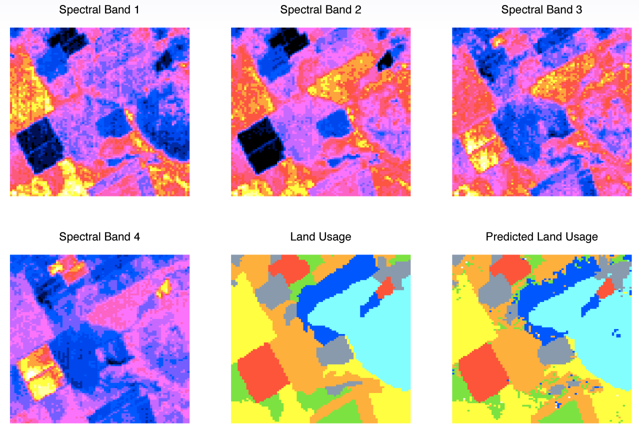


Figure: Usage $\in$ {red soil, cotton, vegetation stubble, mixture, gray soil, damp gray soil}

Example: Relation between salary and demographic variables

- Establish the relationship between salary and demographic variables in population survey data.

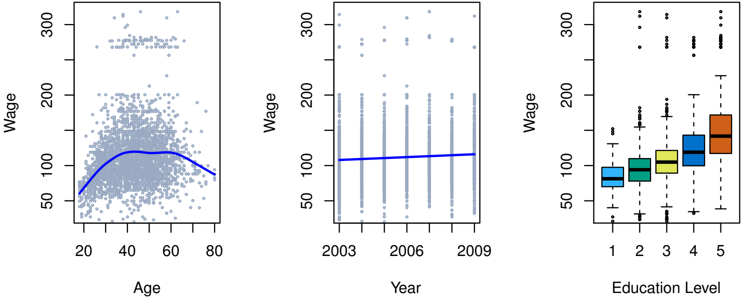


Figure: Income survey data for males from the central Atlantic region of the USA in 2009.

Type of Learning

- Supervised Learning
  - Classification
  - Regression
- Unsupervised Learning
  - Association
  - Clustering
  - Density Estimation
- Reinforcement Learning
  - Agents
- Others
  - Semi-supervised Learning
  - Transfer Learning

References

- The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, Hastie, Tibshirani, and Friedman, Springer.
- In Introduction to Statistical Learning with Application in R, Second Edition, James, Witten, Hastie, and Tibshirani, Springer.

Thank you!