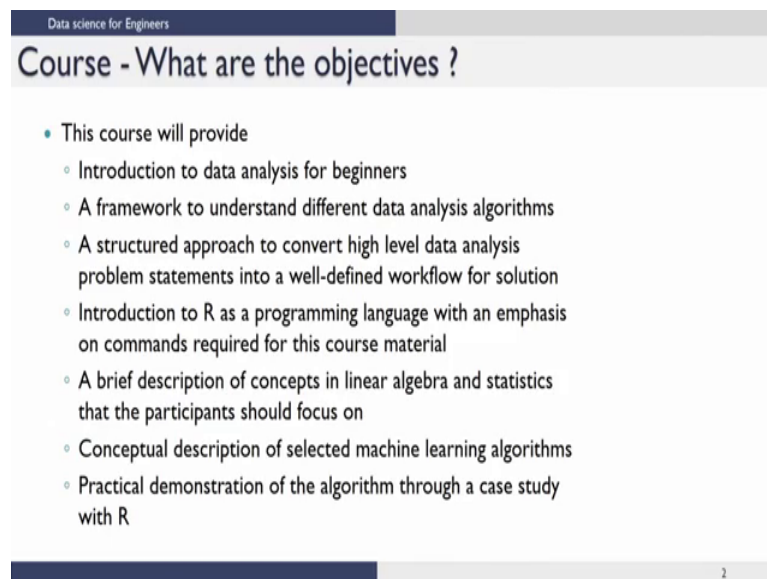**Data Science for Engineers**
Prof. Raghunathan Rengaswamy
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Lecture – 01
Data science for engineers - Course philosophy and expectation

Welcome, to this course on Data Science for Engineers. My name is Raghunathan Rengaswamy. I am a professor in the Indian Institute of Technology at Madras. I will be teaching this course with my colleague professor Shankar Narasimhan also from IIT, Madras. The, teaching assistants for this course are Doctor Hemanth Kumar Thanero and Miss Shweta Shridhar. In this very brief video, I am going to talk about the course philosophy and the expectations that you could have from this course. Let us start with the objectives of the course.

(Refer Slide Time: 00:59)

First off, I want to say, this is the first course on data analysis for beginners. So, this is for people who want to learn data analytics who have not been practicing it for a long time and so on. However, while we say this is a data analysis course for beginners, it would still be a substantial amount of information substantial amount of mathematical concepts and more conceptual ideas that we will have to teach.

So, while it is an introduction course it is still significant amount of effort and learning that that we expect the participants to get out of this course when we talk about data

analytics, there are several algorithms that one could use for doing analytics. So, as part of this course, we will try as much as possible whenever appropriate to explain all the concepts in terms of the data science problems that one might use them to solve. So, in that sense, you try to give you a framework to understand different data analysis problems and algorithms and we will also as much as possible try and provide a structured approach to convert high level data analytic problem statements into what we call as well defined workflow for solutions. So, you take a problem statement and then see how you can break it down into smaller components and solve using an appropriate algorithm.

So, these are at a conceptual level what you would expect the participants to take out of this course. For teaching data analytics or data science it is imperative that you do coding in a particular language there are many possibilities here as far as this course is concerned we are going to use R as a programming language. So, as part of this course R will also be introduced and the emphasis here will be on the aspects of R that are more critical for what you learn in this course. So, in other words commands that are required for this course material will be dealt in sufficient detail.

So, that is as far as a programming language is concerned for learning data science. In terms of the mathematics behind all of this we will describe important concepts in linear algebra that we think or critical for good understanding of machine learning and data science algorithms we will teach those and we will also teach statistics that are relevant for data science. Other than this will also have modules on optimization ideas and optimization that are directly relevant in machine learning algorithms, we will also provide conceptual and descriptions that are easy to understand for selected machine learning algorithms and whenever we teach a machine learning algorithm we will also follow it up with another lecture where the practical implementation of an algorithm for a problem statement is demonstrated and that demonstration would take place and we will use R as the programming platform.

(Refer Slide Time: 04:48)

Course- What it is not?

- This course is not for practitioners of advanced data analysis
- This course is not about big data implementation concepts such as map reduce, hadoop frameworks and so on
- Not an in-depth R programming course
- Only a selected few machine learning techniques that are most relevant for a beginner are taught
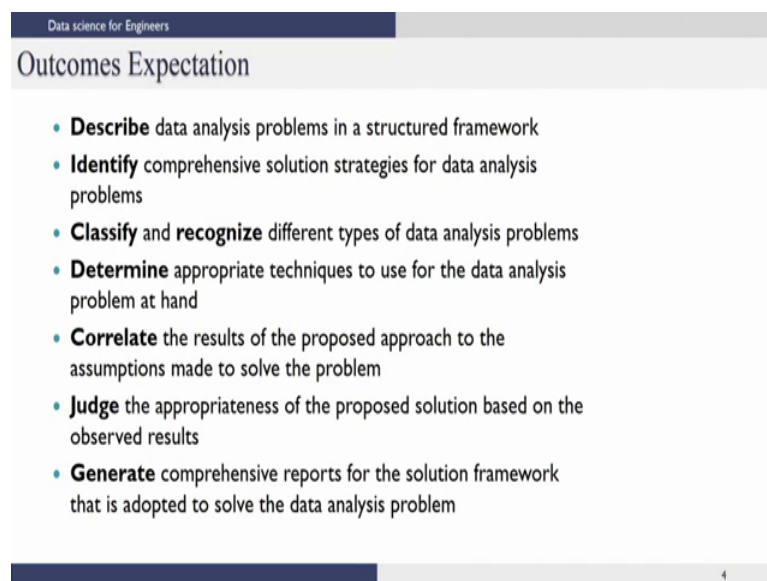
While we talk about what the objectives of this course are it is also a good idea to understand what this course is not about. As I mentioned already if you are a very advanced data analysis practitioner then there are other courses which are at more advanced levels that are relevant, this course is at a basic level for someone to get into this field of data science. We will be teaching a course on machine learning later which might be more appropriate for people of this category. This course is also not about big data per se and we are not going to cover big data concepts such as map reduce, hadoop frameworks and so on.

This course is more about the mathematical side of the data analytics, so, we are going to focus more on the algorithms and what are the fundamental ideas that underlie these algorithms. While we will use R as a programming platform this is not an in depth R programming course where we teach you very sophisticated programming techniques in R the R programming platform will be used in as much as it is important for us to teach the underlying data science algorithms.

Now, there are a wide variety of machine learning techniques there are a number of techniques that could be used and in a nine week course we have to pick the techniques that are most relevant, not only that since we think of this as a first course in data science. Ah, we also have to spend enough time covering the fundamental topics of linear algebra statistics and optimization from a data science perspective. So, that takes quite a few weeks of lecture. So, we are going to pick a few machine techniques which we believe are the most relevant for a beginner.

So, you understand the basic ideas in data science you get a fundamental grounding on the math principles that you need to learn and then you put all of this together in some machine learning technique. So, you understand some machine learning techniques where all of these ideas are used and we have picked these techniques in such a way that you can understand data science better and also use these in some problems that might be of use or interest to you.

(Refer Slide Time: 07:30)

So, in terms of a idea of what outcomes we would expect when a participant finishes this course there are many things that you can do, but these are some categories of skills that that we would expect you to generate. So, you would expect you to be able to describe data analysis problems in a structured framework, once you describe that would expect you to identify some comprehensive solution strategies for the data analysis problems. Ah, classify and recognize different types of data analysis problems and at least to some level determine appropriate techniques.

Now, since we do not teach you wide variety of techniques within the gamut of techniques that you are taught you will be able to identify an appropriate technique that you can use and in this course, we emphasize this important idea of assumption validation. So, you make some assumption support the data that you are dealing with and then those assumptions tell you what algorithms you should use and then once you run the algorithm you get the results and see whether your assumptions are validated and so

on. So, you would be able to think about how you can correlate the results of whatever you have done to the assumptions you made to solve the problem and then see whether that makes sense whether the solution makes sense and so on.

So, that is where we talk about judging the appropriateness of the proposed solution based on the observed results and ultimately ah, we would expect you to be able to generate comprehensive reports on the problems that you solve and then be able to say why you did, what you did, so, that is an important aspect of what we are trying to cover.

So, if you stick with us and get through all the eight weeks of this course and also diligently work on the assignments that are provided at the end of every week then we hope that you learn the fundamentals of data science, you get some fundamental grounding on important ideas and the math that you need to learn to understand data science and take this learning forward in terms of more complicated algorithms and more complicated data science problems that you might want to solve in the future.

So, I hope all of you learn and enjoy from this course and we will see you as the course progresses.