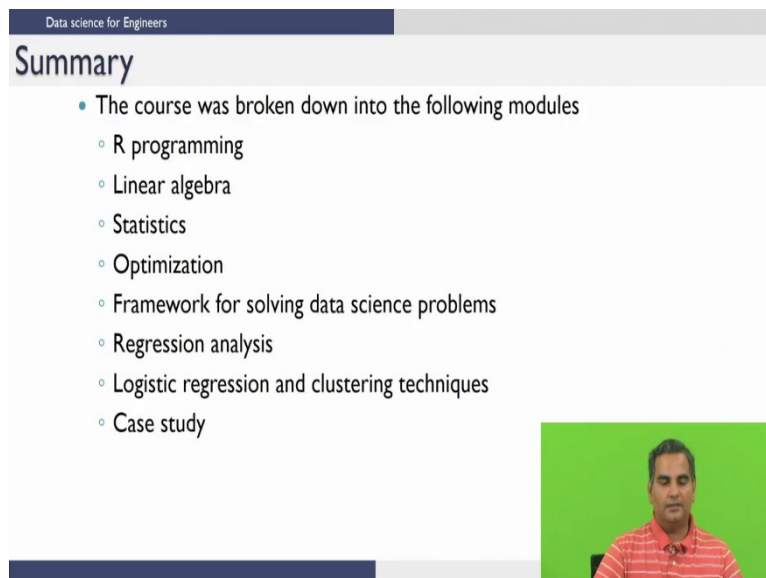


Data science for Engineers
Prof. Raghunathan Rengaswamy
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Lecture - 50
Data science for engineers – Summary

So, with this we come to the end of the course on Data Science for Engineers. I am going to do a quick summary of the course. I hope all of you had a productive time looking through the videos and learning interesting ideas in data science.

(Refer Slide Time: 00:34)



Data science for Engineers

Summary

- The course was broken down into the following modules
 - R programming
 - Linear algebra
 - Statistics
 - Optimization
 - Framework for solving data science problems
 - Regression analysis
 - Logistic regression and clustering techniques
 - Case study

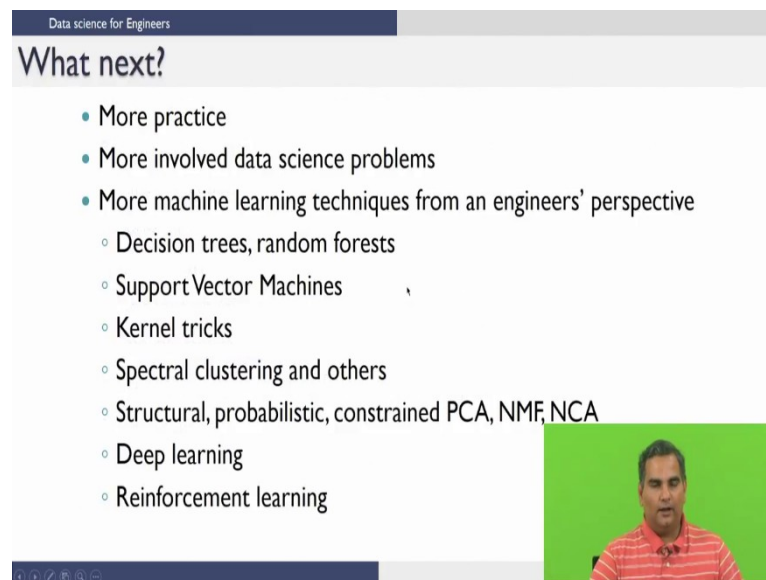
In summary we broke down this course into the following modules. We started with R programming as the programming platform for solving data science problems. So, we had a set of lectures explaining the important concepts of R programming that are useful from a data science viewpoint.

We followed that up with lectures on fundamental ideas in linear algebra and statistics that are useful for data science and machine learning. We then introduced the idea of optimization and optimization algorithms that are useful for us to understand machine learning algorithms and make sense out of why we get some results when we run certain machine learning algorithms and so on.

After that we described a framework for solving data science problems. This is a framework that you can use to conceptually break down a large data science problems into smaller sub problems in some work ow fashion. We then moved on to describing regression analysis where we described techniques for univariate and multivariate problems.

Here we focused on building appropriate models identifying the goodness of models and so on. Then we looked at classification techniques logistics, regression and kNN and then we also described clustering techniques such as k-means clustering. And as part of this course there is a case study for you to practice which will be available on the website.

(Refer Slide Time: 02:27)



The slide is titled "What next?" and is part of a presentation for "Data science for Engineers". It contains a bulleted list of topics to be covered next:

- More practice
- More involved data science problems
- More machine learning techniques from an engineers' perspective
 - Decision trees, random forests
 - Support Vector Machines
 - Kernel tricks
 - Spectral clustering and others
 - Structural, probabilistic, constrained PCA, NMF, NCA
 - Deep learning
 - Reinforcement learning

In the bottom right corner, there is a small video inset showing a man in a red and white striped shirt speaking against a green background.

Now, that is you have done this course and hopefully you have learned enough from this course. What is the logical next step if you were excited by this course and want to know more about data science after doing this course? I would say first is to do more practice on the same ideas that I have been taught in this course. So, you might want to look at other problems and other practice examples and exercises for the concepts that we have already taught in this course. So, that is the first thing to do.

Once you do that the data science problems that we described in this course are rather simple. So, you might want to look at how people solve more involved data science problems and whether you can use the framework to break them down into smaller problems and then see whether you can learn more about these problems. Now, as I mentioned before we teach only very few machine learning techniques

in this course for the beginners. However, there are many many more algorithms that are out there such as decision trees, random forests, support vector machines, kernel tricks and so on. So, we have listed many of the commonly known and used algorithms for more complicated or more complex machine learning problems.

So, the next logical step would be once you master the material that has been taught in this course is to look at learning these algorithms and you could use the same notion of understanding the assumptions that are underlying these algorithms to get a good idea of why these algorithms work the way they work. And also would understand the technical details of these algorithms in terms of what is the learning rule and why does it work and so on.

So, a data science is a very vast field. Obviously, there is a lot of interest this lot of buzz around this field and it is only going to get even more important as we go along because more data is going to be generated and one would expect the computational power to increase even more for the next few years. And the math and the algorithmic details are being proved by many researchers, so one would expect even better algorithms to come down in the next few years.

So, a mix of all of these factors make it important that everyone have some fundamental understanding of data science and people who are really interested in this field have much more in depth understanding of techniques such as the ones that are listed here and so that you are able to solve much more complicated problems which have much more value in real life. So, happy learning and hope this course was useful.

Thank you.