We have already seen the typical AIML pipeline of data/problem to representations and then to understanding/classification. In the session 3, we also had a closer look at the problem of perception, and how the area is developing.

We also got introduced to the notion of deep learning and neural neetworks.

We now know the three important steps are:

1. Feature extraction/representation
2. Classification
3. Performance analysis.

We are exposed to all the three. Good start. The first step, where we want to build a representation for the data/observations has been dominated by the intuitive and some time hand crafted methods till recently. This has been a concern for many of us also what is a good representation for my problem? In recent years, statistical methods as well as learning based methods have become very important in this space. In this lecture we look at this module more closely.

We primarily look at three different types of data today:

1. **Text** which we see in many domains like emails, web pages, documents, tweets, product descriptions etc.
2. **Images** whcih we see in a number of situations like videos, biometrics, remote-sensing, survillience, autonomous navigation etc. and
3. **Speech and Signals** which is a 1D data with applications in speech recogni-tion, assistive systems etc.

All of them have rich history in representation. They are on the center stage now. We may be carrying AIML algorithms that classify these modalities in our own personal mobile phones. These are the areas where AIML had made the biggest impact in the recent years. In this lecture, we will look at two representations (one pre-deep learning era (say classical) and the other which use the modern neural networks (say modern)). In the associated laboraties we will follow up with experiments which will expose these representations and the utilities. We also see a number of other data descriptions which has a combination of number, text, image, speech etc. (eg. product details). It is worth thinking, how can we now learn representations that are useful for such data. We hint at them too in this lecture. In our previous lectures, we were motivated by the metric/geometric view

points of the data. Apples and organges come together in some 2D space. How- ever, we always wondered what about categorical variables? Text/words is an

example of such representations. Is there some notion of distance/similarity

between words? Can we use our favourite Euclidean distance on such represen- tations?

An interesting (if you have not seen earlier, suprising) representation of the word is now using word2vec. This representation allows us to do some simple arithmatics like

$man - woman + queen = king$

or

$walking - swimming + swam = walked$

Yes, it is not easy to do this additions and subtractions and make sense with our classical representation of the text/word. What we now have is an embedding/representation that we can manipulate like a simple real vector as shown above. This makes the development of the algorithms easy, and more meaningful. Such learned representations are now becoming more and more pwoerful in many areas.