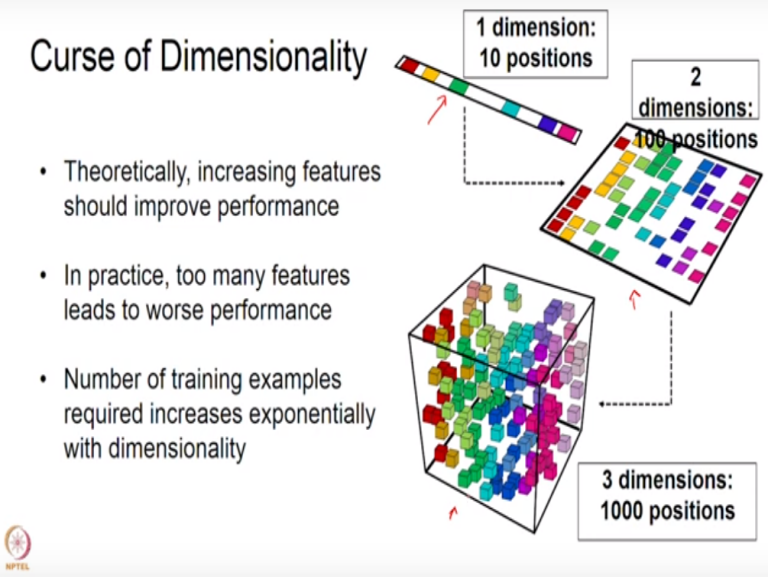
Machine Learning for Engineering and Science Applications
Professor Dr. Ganapathy Krishnamurthi
Department of Engineering Design
Indian Institute of Technology, Madras
PCA – Part 1

Hello and welcome back, so in this video we will look at principle component analysis I promise that we will just take a look at it because this is one of the techniques used for data pre-processing or data normalization when before you use them as input to machine learning algorithms or deep learning algorithms in general.

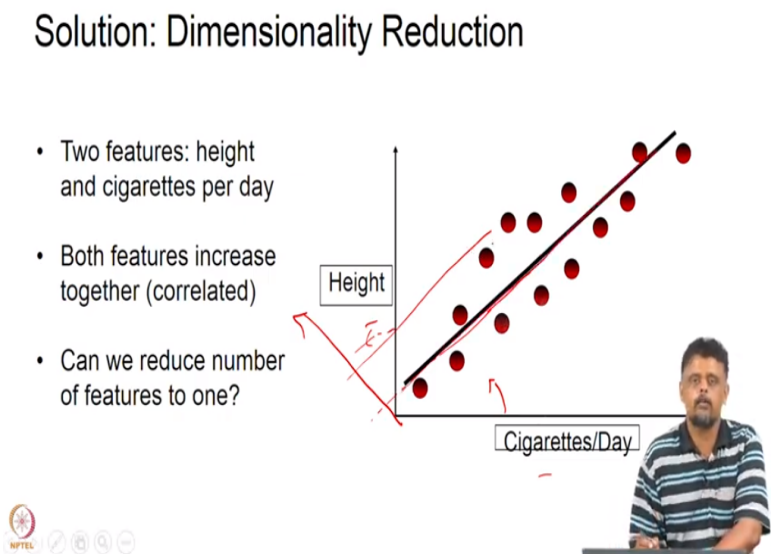
(Refer Slide Time: 00:28)



See theoretically what happens is increasing the number of feature should improve performance however as you increase the number of features there are some problems because as you it turns out that many features leads to worse performance, so if you have lots of feature where leads to worse performance because what happens is we need more training examples if you have more features thinks of it as no filling up and all this is case as this example slow you know if you are a way of only one dimension.

So we can fill up we can sample that dimension with fewer data points however as the dimension becomes two you need more data points to sample the entire space of you problem and also as you as the number of dimension keep increasing the number of points you need in order to sample the space of a problem begins to increase and so if you have lot more dimension and less number of data examples you run into problems with training your algorithm.

(Refer Slide Time: 01:27)



So the solution to this problem is dimensionality desire reduction wherein we reduce the number of features that represents the data, so how do we do this the idea is to reduce the dimension by selection subsets, subset which is by feature elimination and we do that the algorithm does that we are going to see it is what we refer to as a principal component analysis, so in this case we have a data with two feature number of cigarettes per day and height this is some sample of the population and we do not write this point I do not worry about what this data was collected for what are the classification or regression algorithm we just know that there is some these or these two data sets are there.

So there are two features and it seems like what is more important observe here that the feature increase together so they are correlated, so I would like you to go back and think about the naive bayes algorithm where we saw that our, where the where we assumption is that the features are not correlated actually, so in this case if the features are correlated can we reduce the number of features to one so this is a very easier problem to visualize, so that is way we brought with two features so typically in machine learning problems you realize that will be hundreds of features and then you will be forced to deal with the completely differentiate.

So in this case we have a features which is height and cigarettes per day and they simply correlate it also remember if there only two or three feature we can actually plot this correction plots and actually eliminate some of them right that is a possibility but then let us say if you

hundreds of features you can imagine the number of correlation plots you have to do to see how things are correlated or which of the features are correlated actually PCA is another way of looking at tells you which are the most uncorrelated features and which ones are only significant because they are correlated to other features in some way or the other.

So way to do that can we reduce the number of features to one, so the way to that is if we can fit this line you can figure out this line, and we can if you can project the values curl of these training points that line then all we need is the location of this projection along this line that will be your new axis you can think of it. So the way to think of PCS at least in physical problem is the relation of your axis, so if you go back, so this is your axis you have cigarettes per day and height and the idea is now we know that there is some correlation.

So can we rotate this axis these two axis, so this axis come here comes here and this axis rotates this way where the other axis the values along the other axis are very small values are very small and the values along this axis are very large, so then we can afford to ignore the other one.