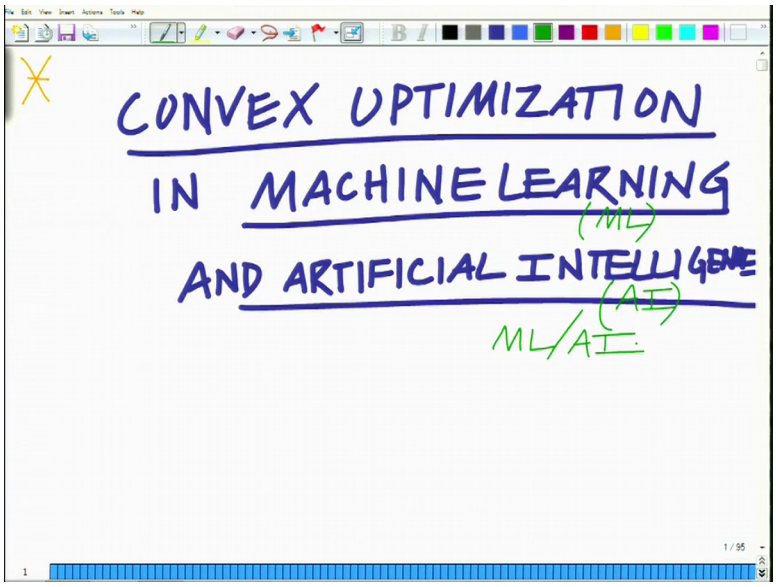
Applied Optimization for Wireless, Machine Learning, Big Data

Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

Lecture - 60
Practical Application of Machine Learning and Artificial Intelligence: Linear Classification, Overview and Motivation

Hello, welcome to another module in this massive open online course. So, we are look in various aspects of optimization. Let us look at another important application of convex optimization and that is in the latest evolving field of machine learning and artificial intelligence. And in fact the applications here are very interesting and have a lot of relevance especially for modern applications all right.

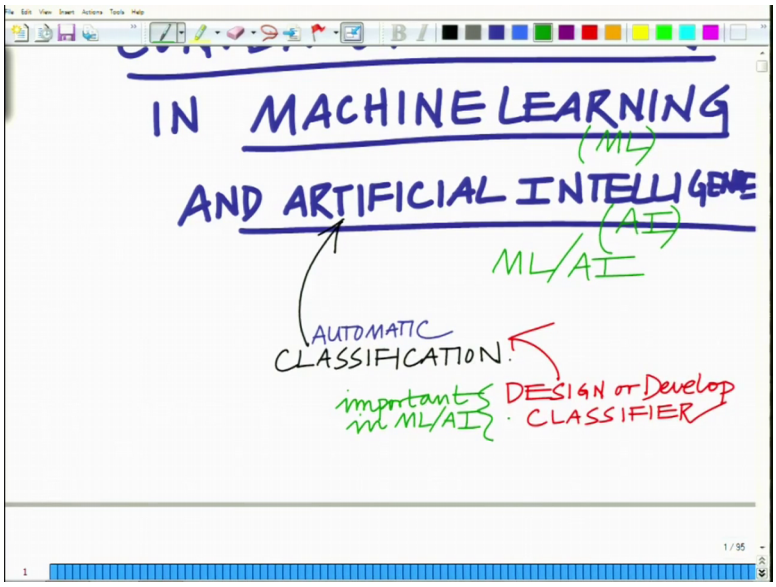
(Refer Slide Time: 00:39)



So, what you want to look at is you want to focus on in for this module, and some other next modules. You want to focus on convex optimization and demonstrate the important role it has to play in machine learning which are currently very hot topics of both research and also application; machine learning and artificial intelligence. And well these are abbreviated frequently for instance machine learning. This is abbreviated as ML not to be confused with another ML that is a maximum likelihood estimate and artificial intelligence as AI.

So, frequently you will see the nomenclature ML slash AI used which implies a machine learning or artificial intelligence applications. And of course this is a huge field. If you look at machine learning or artificial intelligence, there are large number of problems with several several interesting applications. In this particular MOOC, to limit the scope little bit and to focus on some of the basic, yet very relevant problems. We will limit or discussion to the development of classifiers.

(Refer Slide Time: 02:17)



So, we are going to focus on the problem of classification mode think. And these problems can be generalized, in fact lay the foundation for the development of very complicated and sophisticated machine learning algorithm, so classification or automatic classification rather not manual classification not human assisted, but automatic classification.

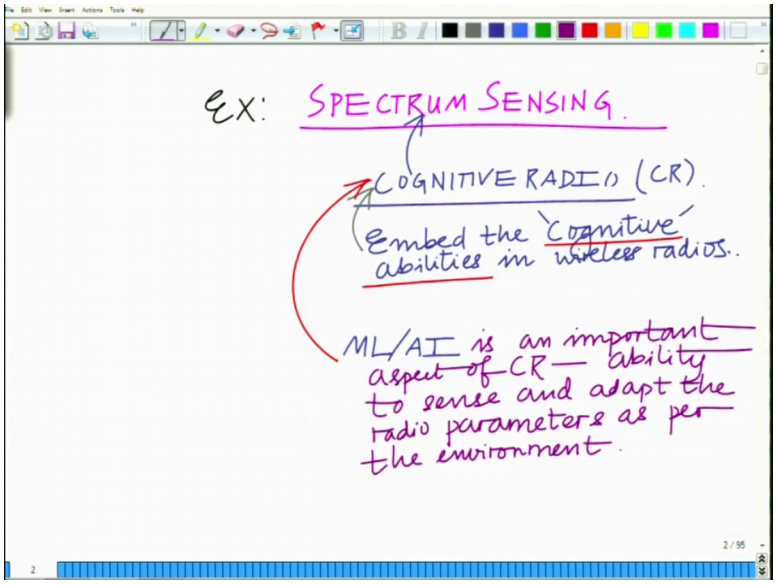
And for this purpose, we have to design or develop or basically we have to develop what is known as a classifier. And this is an important problem in machine learning or artificial intelligence that is to automatically classify alright recognize patterns from data to automatically classify a set of objects belonging to either two different sets or multiple different sets all right.

So, classification; that is classifying various objects various things is one of the most important and fundamental projects or problems in machine learning or artificial intelligence which says the foundation for several advanced problems all right. And, now

for instance this can be for instance let us say you have a video or you have an image from that video you are trying to gather information about the traffic, you tried to classify the number of you will have various objects you client to classify the objects into let us a car versus trucks versus two wheelers and so on.

So, these are all so or you have basically a video grab of a busy area, and you trying to classify for instance the people there belonging to different sets for instance just men, women and so on. So all these are very interesting examples and there is plenty of such examples of classification for this to better illustrated.

(Refer Slide Time: 04:22)



Let us look at a typical example in modern wireless communication and that is of spectrum sensing. So, the example like this has also so spectrum sensing let us take an example from spectrum sensing. And the spectrum sensing problem arises in one of the latest wireless technologies which is known as cognitive radio. We might have already seen this to some extent, but no harm in repeating in cognitive radio or C R.

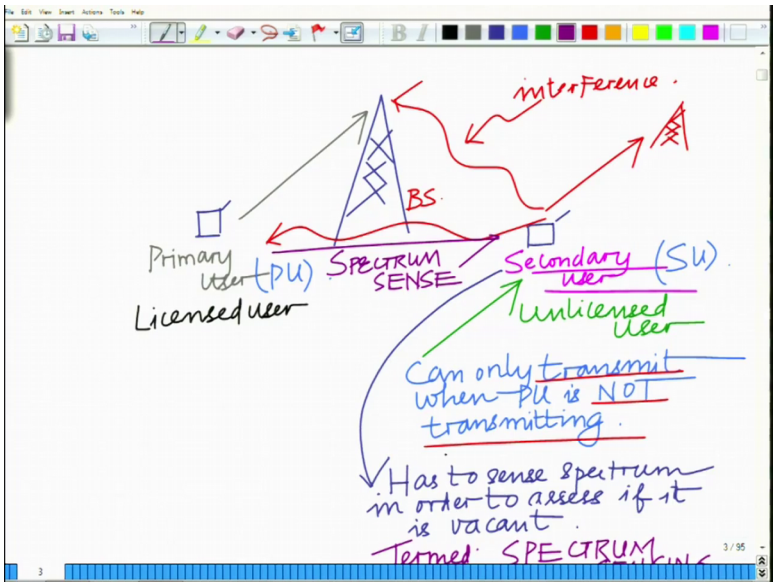
What happens in a cognitive radio technology is that the wireless device is embedded with intelligence, and therefore ML or AIs an important aspect. So, in cognitive radio cognition is basically is similar to what the human mind does is extract patterns or recognized rights. So, cognition is an important aspect of, you say a cognition is important aspect of the human brain.

And the idea is to embed cognition or embed this kind of intelligence alright. The cognition ability cognitive ability similar to the human brain in radios so, cognitive radio is to embed the ability of cognition or embed cognitive abilities in wireless devices or wireless radios ok, so that is the basic idea to mimic some of the functionalities of the brain in a wireless communication system.

And therefore, ML or machine learning or artificial intelligence is an important aspect that is AI is an important aspect. The ability to sense and adapt of C R basically this is nothing but the ability to sense and adapt the radio parameters as per the environment, in the sense. The radio environment what we mean by that is the signal level, the noise level, the difference level, the kind of standard or the kind of technology that is being used is CDMA, OFDM or is it Wi-Fi.

So, the ability for the radio to sense, sense the environment is that interference is that too much interference was the channel vacant. So, the ability for a radio to sense its environment and adapt itself is a very important aspect of cognitive radio. In fact, it is a sequence side the central aspect of cognitive radio. And therefore, ML machine learning or artificial intelligence which basically is concerned with basically or basically concerned with extracting learning rules right based on sets of data or let so learning patterns recognizing pattern based on sets of data. And in using them later with a very high degree of pretty is an important aspect of cognitive radio alright.

(Refer Slide Time: 08:24)



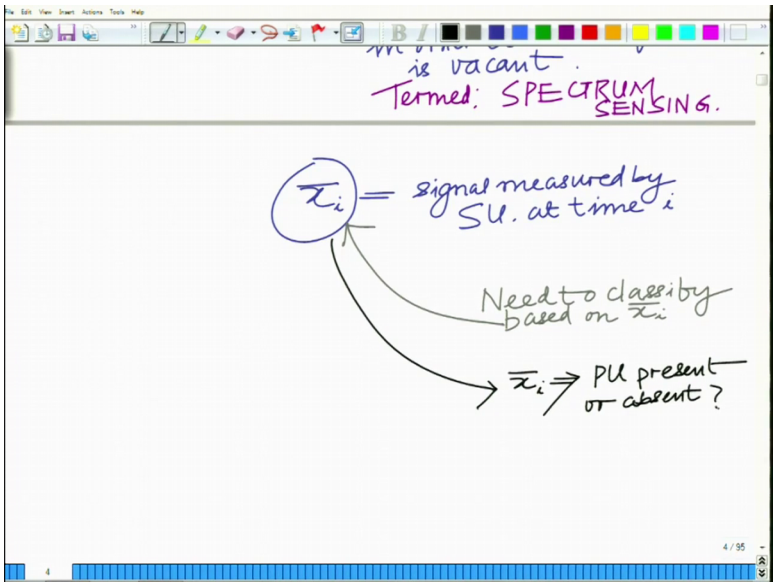
For instance, a simple problem in cognitive radio can be the following and this a very one of the most basic problems in cognitive radio, which is the following that is when you have a primary user or the user who is licensed to transmit in a certain spectrum, so you have primary user or basically let us say a licensed user.

And in addition you have a let us say secondary user. Now, this is the unlicensed user can be communicating with the different base station that is the secondary a different base station ok, but remember causes interference at the primary base station. So, this is base station this is; indirect causes interference also at the primary user ok. So, this is your secondary user.

So, naturally the secondary user can transmit, also the secondary user alright causes interference to the primary user, because if he or she the second user wishes to access the same band all right, but only when it is vacant which implies, because if the primary user is transmitting and the secondary user who is unlicensed user ok, the secondary user is also the unlicensed all right. It is an unlicensed user; the unlicensed user and if the secondary user causes interference to the primary user, therefore secondary user can only transmit, when the primary user is not transmitting.

When the primary user what P U, secondary user is a S U. When the primary user P U is not, can only transmit when the primary user is not transmitting and or the P U is absent, so as to say alright. And therefore the S U, so the S U has to sense alright to sense the environment or has to sense the spectrum to assess, if the spectrum is vacant, alright this process is termed the spectrum sensing all right. So, the secondary user in order to assess if it is vacant.

(Refer Slide Time: 11:54)

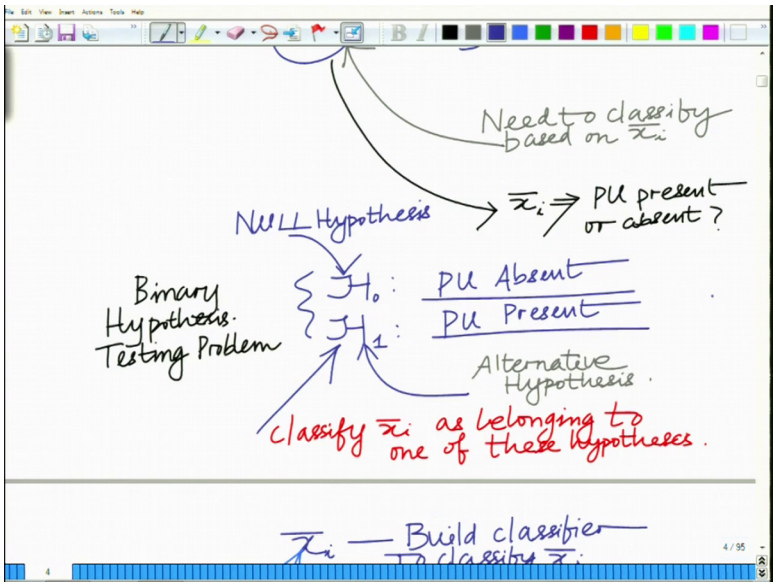


And this is termed as; this process is termed as spectrum sensing that is sensing the radio spectrum. This is an important part and this is important aspect of cognitive radio in any sense, because the cognitive radio has to sense the wireless environment and adapt the radio process all right. So, this has to sense the spectrum all right.

And naturally this spectrum specters sensing process at the secondary user is going to be based on some measurements that are done by the second user of the environment alright based on the signal let us say x bar that is sensed by the secondary user all right. So, let us say x bar i this is the signal measured by secondary user at time i that is we denoting by x bar i. Now, we need to classify now what we now when we say classify, now what is the relevance, ML need to classify based on this x bar i alright.

So, what we are saying is depending on the measurement we have to assess, if this correspond this x bar i corresponds to the primary user being present or the primary user being absent that is a fundamental spectrum sensing problem. So, x bar i classify in the sense x bar i implies P U present or P U present or absent.

(Refer Slide Time: 14:17)

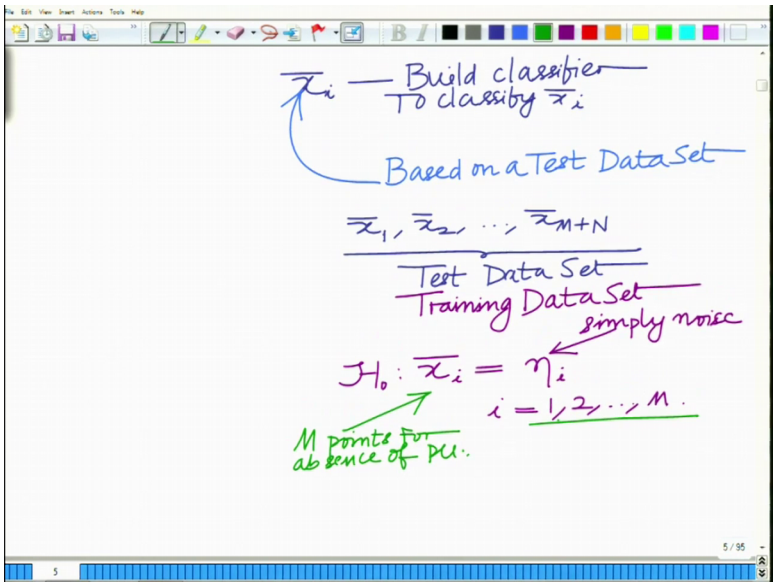


Now, how do you classify this, let us build a framework for this. Let us consider to hypothesis we can only hypothesis all right. Hypothesis is known as hypothesis 0 that is primary user and then we have a hypothesis 1 0 means primary user is present. There are two hypotheses either primary user is present or primary user is absent, this is a hypothesis this is term as the hypothesis 0 or this is the NULL hypothesis. This is termed as the alternative hypothesis.

Null hypothesis we have the alternative hypothesis. And we have to assess if based on the current signal, where which of the hypothesis is 2 alright. So, there are two hypotheses the NULL hypothesis, the alternative hypothesis and based on the signal measurement we have to assess which of the hypothesis 2. And this is known as a binary hypothesis testing problem ok. So, this is known as a binary hypothesis testing problem. This is standard problem in detection this is a binary hypothesis in testing problem.

And we have to classify x bar i the measurement as belonging to one of these hypothesis. So, the idea is classify x bar i as classify x bar i as belonging to one of this hypothesis ok, so that is basically your classification problem or your binary hypothesis testing problem. Now, towards this end we have to build a classifier. Let us with classify the data to classify this measurement x i, we have to build a classifier. And we will build that classifier initially on the base of some data that is available some data set or test data set that is available with us all right.

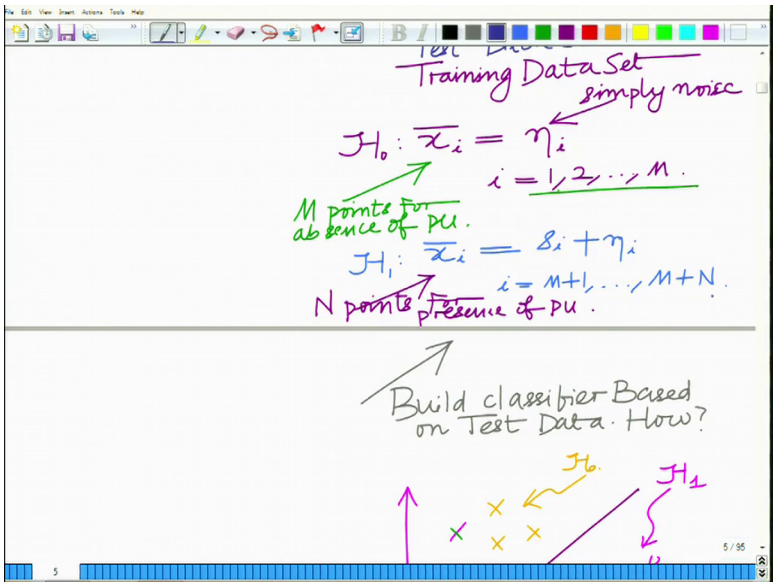
(Refer Slide Time: 17:00)



So, basically to classify x bar i we have to build a classifier classify x bar i. And this building of classifier based on a test data set. Let us, consider the test data x bar let say M plus N I will explain this soon x plus, this is our test data set all right. You can also this is a training data to build a classifier to build any classifier in any classification problem you need a training data set.

Let us say in this training data set we have points x bar i equal simply the noise that is no primary user hypothesis 0. If there is no noise and i equals 1, 2, up to M all right. These are M points which correspond to i equal to 1 to M. So, M points for absence of primary user. The M points, this is the M points for the absence of primary user.

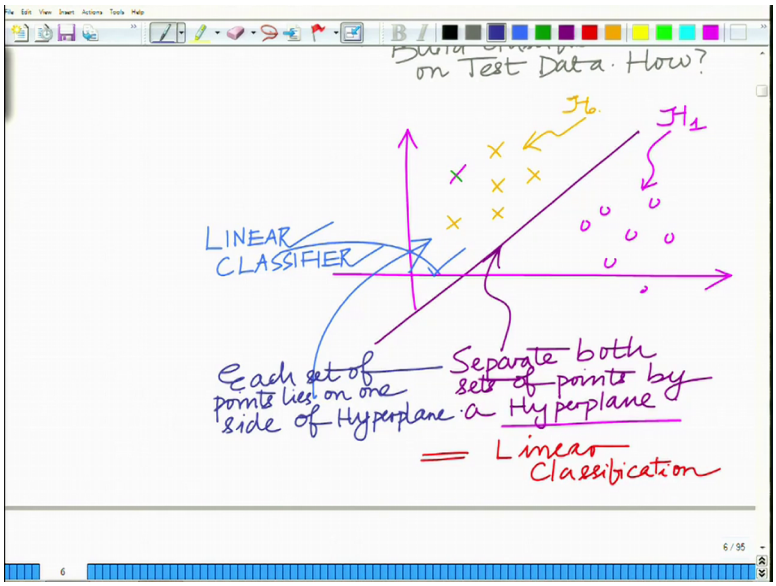
(Refer Slide Time: 19:07)



And let us say we have the rest the M points. Remember our test data is M of size M plus N, so x bar i equals signal plus noise s i plus n i or sequence side plus eta i; i equals M plus 1 up to M plus N ok. And these are N points for the presence of primary in the test data set ok. And we have to build.

So, we have M plus N point so test data is basically you have M plus your test data set x bar x bar 1 x bar 2 up to x bar M plus N all right, test data set M plus N point out of this M points corresponds to absence of the primary user, the null hypothesis, N points correspond to the presence of the primary user or the alternative hypothesis. And we have to build our classifier based on this test data set.

(Refer Slide Time: 20:55)



And now the we have to build our classifier based on the test data set, the question is how and how do we build it? For instance let us take a simple example, let us plotted let us say we have a 2 dimensional plot ok and let us say this data corresponds to hypothesis h 0. We expect to see this kind of clustering behavior this point corresponds to hypothesis 0 or close to each other points correspond to hypothesis 1 or close to each other alright. Because it would be very weird, if all of them are occur in a single cluster, then would be difficult to classify or separate them all right.

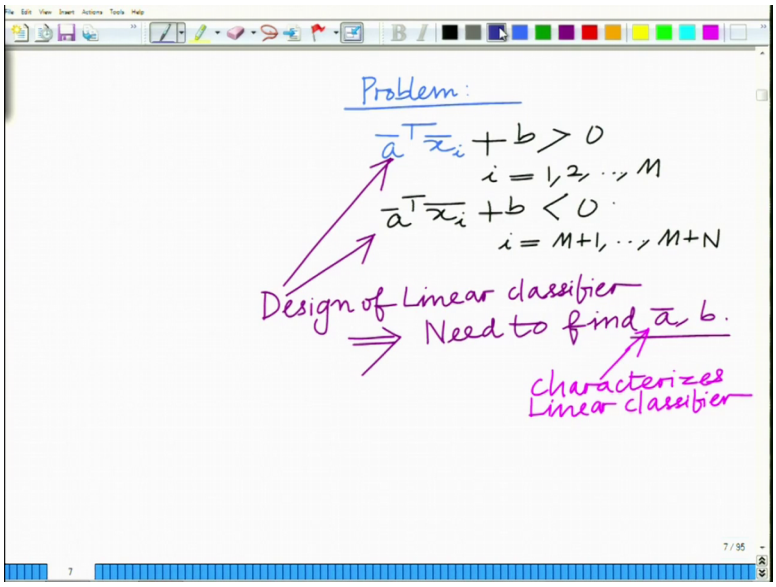
So, there as you expect to see some logical separation between them and these points correspond to hypothesis H 1. And now what we want to do is you want to classify these two uninteresting thing could be to separate these by a hyper plane. Separate data sets of this both; separate both sets of points by hyper plane. And this is basically termed as linear separation or linear classification. This is termed as linear classification. This is known as the linear hyper plane which separating this is known as a linear classifier or a linear discriminant. One simplest classifiers, yet very powerful or to you are saying is very simple you have two sets of points all right belonging to the two hypothesis these are drawn from the training data alright.

And we want to build a classifier all right. What is our classify? The classifier is the hyper plane and the set of points belonging to each hypothesis lies on different sides right. The sets of points belong to each like the one set lies on one side the hyper plane

and other set lies on the other side of the hyper plane. And there by this hyper plane it basically separating both the sets of points belonging to the two hypothesis and that is how we are classifying these sets of points alright.

So, the problem can be represented as follows. So, each set of points lies on one side of the hyper plane that is how we are classifying. Each set of point lies on one side of hyper plane all right, all h 0 points on one side, and all h 1 point on that side.

(Refer Slide Time: 24:22)



So, and this problem can be formulated as follows, remember you can formulate this as a bar transpose, because hyper plane is a bar transpose x bar plus b, so a bar transpose x bar i plus b this is greater than 0, let say for i equal to 1, 2 up to M. And a bar transpose x i plus b less than 0 for i equals M plus 1 up to M plus N. Remember the hyper plane divide this space into two parts, points on one side of hyper plane are a bar transpose x bar plus b greater than 0, points on the other side of the hyper plane satisfy a bar transpose x bar plus b less than 0 all right.

And therefore naturally if the set of, the set of signals of the setup that the set of points in the test set corresponding to x bar 1 x bar to up to x bar m, alright satisfy the equation a bar transpose x bar x bar transpose x bar i plus b greater than 0, and other set that is i equal to M plus 1 up to M plus N satisfy a bar transpose x bar i plus b less than 0. Naturally, these lie on both these sets lie on different sides of hyper plane and therefore we would have achieved classification that is classifying these.

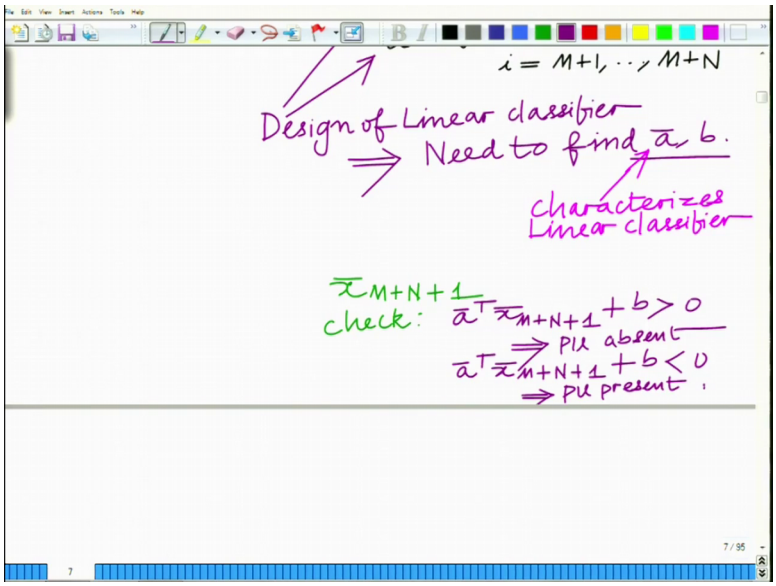
And what is meant by finding such a classifier? Finding such a classifier is nothing but finding the hyper plane which basically nothing but the hyper plane is characterized by these parameters a bar and b. Therefore, it intern implies that we have to we have to estimate or we have to compute somehow come up with estimates for a bar and b all right. So, design of classifier or design of linear classifier implies need to find a bar comma b. This characterizes the hyper plane or this characterizes the classifier this characterizes the linear classifier.

So, basically is very interesting what we have done is, we have boiled down this machine learning or artificial intelligence problem of classification. In this cognitive radio system classifying the set of points corresponding to the absence of the primary user, and presence of the primary user belong to the test set into the design of a hyper plane alright. Or into finding a hyper plane, which achieve this separation such that all the sets all the points belong into the presence of the primary user, lie on one side of the hyper plane, all the points belong to the absence of the primary user lie on the other side of the hyper plane.

So, if you find these parameters a bar and b corresponding to such hyper plane, then we would have build this classifier. And this classifier can then subsequently be used to basically classify the further things. So, let us say you make a measurement x bar i at time M plus N plus 1. Now, you can naturally use this classifier put it into the equation a bar transpose x bar plus b. See if it is greater than 0 it is greater than 0, it means the primary user absent. If it is less than 0, it implies primary user is present.

So, if I remember correctly i equal to 1 to M is P u is I am sorry this is hypothesis h 0, which P U is absent and hypothesis h 1 which is P u is present. And so we have this data set so 0 1 x bar 1 to up to M. These are 4 primary user these are 4 primary user absent i equal to M plus 1 up to N plus 1. These are points for primary user present ok.

(Refer Slide Time: 29:08)



And now if you have point at time M plus N plus 1, you can check. So, if you have x bar or if you would have time, M plus N plus 1. Once we have the classifier check, if a bar transpose x bar M plus N plus 1 plus b greater than 0, this implies P u is absent, or a bar transpose x bar M plus N plus 1 plus b less than 0, this implies the P u is present ok.

So, thereby you build a classifier and then you can subsequently you build a classifier based on the training data. And then you can subsequently use the classifier to practically classify the measure signal, and to eventually sense the spectrum that is preside, if the primary user is present, or if the primary user is absent all right. So, we will stop here, and we will see how convex optimization helps in building the classifier in the subsequent modules.

Thank you very much.