Essential Mathematics for Machine Learning

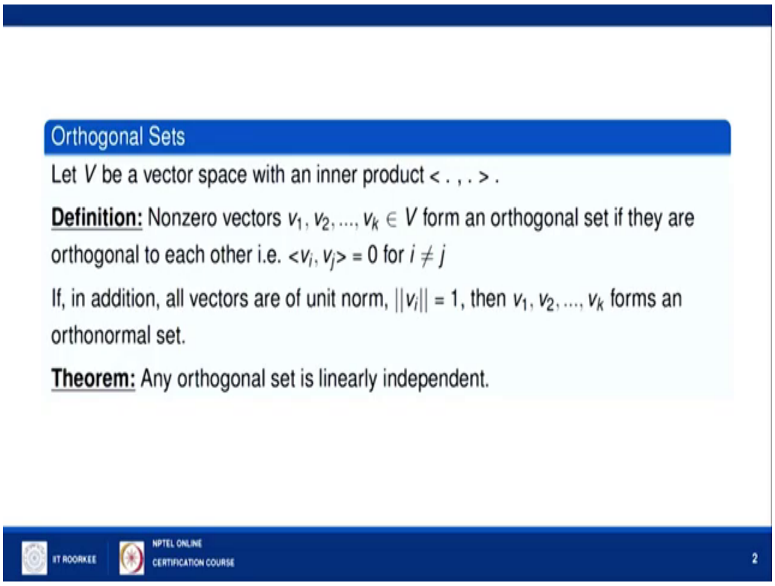
Prof. Sanjeev Kumar
Department of Mathematics
Indian Institute of Technology, Roorkee

Lecture - 14
Gram Schmidt Process

Hello friends. Welcome to lecture number 14 of this course on Essential Mathematics for Machine Learning. Today, we will discuss about Gram Schmidt Process. It is a process from which we can transform a set of li vectors to a set of orthogonal vectors or in fact, orthonormal vectors.

Because, you know in many basis we require orthonormal basis and therefore, this process is really very important, when you need those kind of orthonormal or orthogonal basis because, you can transform any basis to orthogonal or orthonormal using this process.

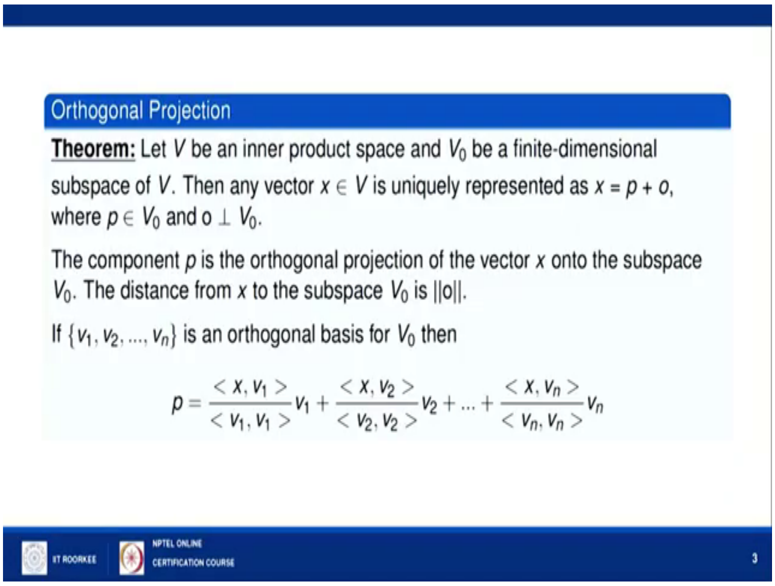
(Refer Slide Time: 01:06)



So, first let me define the orthogonal sets. However, many times we have gone through this definition, but just to recall you that let be V be a vector space with an inner product. So, basically V is an inner product space here nonzero vectors v 1, v 2, v k. So, these are k vectors from the inner product space capital V. They, form an orthogonal set if they are orthogonal to each other.

It means, the inner product of v i v j equals to 0, whenever i not equals to j. If in addition all vectors are of unit norm that is the inner product of a vector v i with itself equals to 1, for all i 1 to k. Then it is called an orthonormal set, means v 1, v 2, v k are forming an orthonormal set. Again if you are having orthogonal set, then it will be linearly independent, but reverse is or converse is not true. So, if you are having a set of k vectors that is orthogonal; obviously, those k vectors will be linearly independent.

(Refer Slide Time: 02:20)

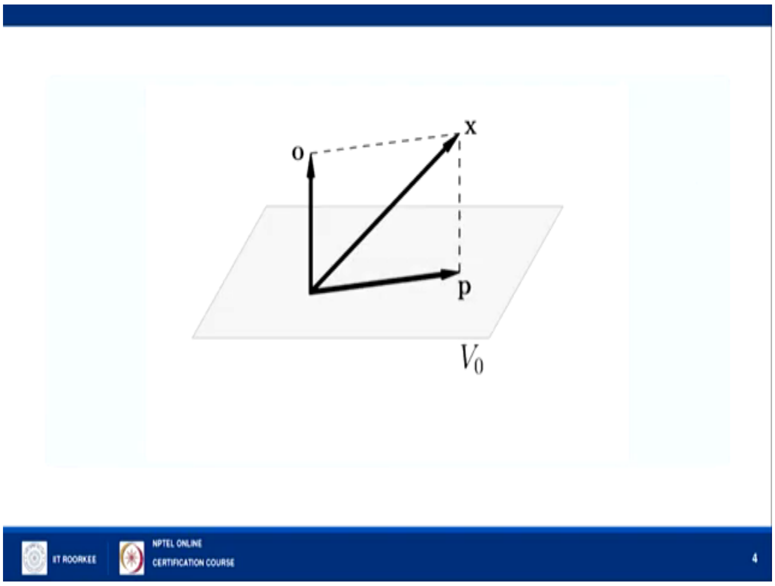


Now, we you have seen this theorem already in orthogonal projection, but we have to recall it in a wider setting means, where we are talking about a subspace. So, let V be an inner product space and V 0 be a finite dimensional sub space of V. So, what V is an inner product space and V naught is a finite dimensional sub space of V.

Then any vector X belongs to V is uniquely represented as X equals to p plus o, where p is a vector in V 0 and o is a vector which is perpendicular to V 0 that is V 0 complement. The component p is the orthogonal projection of vector X onto the subspace V 0. The distance from X to the subspace V 0 is given by the length of this vector o because, it is a perpendicular distance.

So obviously, it will be distance from X to the subspace V naught. If v 1, v 2, v n is an orthogonal basis for v naught, then you can calculate very easily p. So, p will become X v 1 and the inner product of X with v 1 upon v 1 norm multiplied with vector v 1, then inner product of X v 2. So, what it is giving you the component of the vector X in V naught subspace.

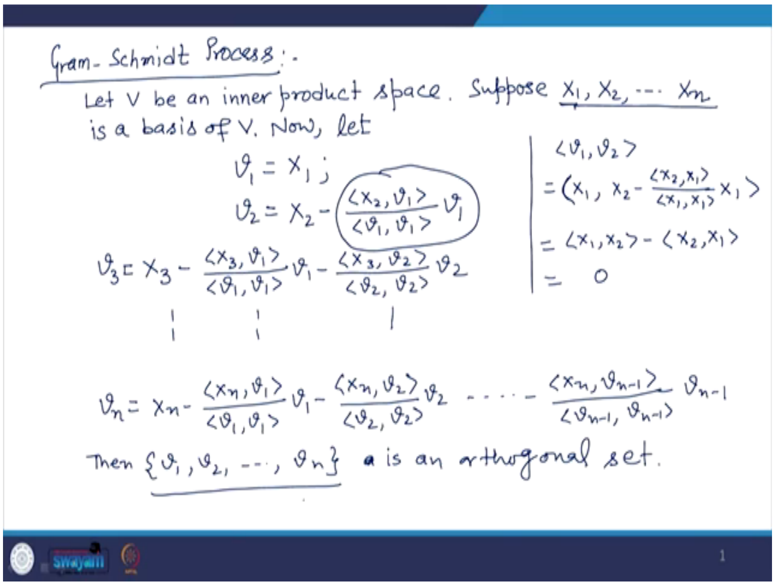
(Refer Slide Time: 04:03)



So, you can see by this graphic, so, you are having this vector X and this is the subspace V naught. So, what is p? So, p is the component of X in subspace V naught and o is the perpendicular component. So, I can write this X as p plus o and that I can write it uniquely, where p is a vector in V naught as you can see here in this V naught subspace and o is perpendicular to p or even all the vectors lie in the subspace V naught.

And hence distance of X can be given by this and that is the length of the vector o. So, these are some results which we have seen earlier also in the orthogonal projection lecture. Now, we will use these results to find out a process by which we can convert a set of linearly independent vectors to a set of orthogonal vectors. And that process is called Gram Schmidt process.

(Refer Slide Time: 05:15)



So, let V be an inner product space. Suppose, X 1, X 2, X n is a basis of V. So, these are the vectors in V and they form a basis of V. So, better I write it in this way; so, X 1, X 2 and X n since they are vectors.

So, now since X 1, X 2, X n are basis of V, so, they are linearly independent. Now, I want to convert the set of vectors X containing the vectors X 1, X 2, X n into a set of vectors those are orthogonal. So, now let v 1 equals to X 1. So, whatever first vector you are having here that I have written as such. Now, I want to make second vector v 2 and I want that this vector v 2 should be orthogonal to v 1 that is your X 1.

So, for making it orthogonal what I need? I need to separate the component of in the direction of X 1 from the vector X 2. Means whatever component of X 2 is there in the direction of X 1, if I subtract that from X 2 then what will happen? After subtraction whatever vector you are getting that will be perpendicular or orthogonal to X 1. So, I will use this concept only. So, v 2 equals to X 2 and now from X 2, I will subtract the component of X 2 that is in direction of X 1.

So, how I will do it? I will write like this. X 2 v 1 upon v 1 v 1 means, I am subtracting in the direction of X 1 that is your v 1 only into so, in the direction of v 1 so into v 1. So, if you notice this is the component of X 2 in the direction of v 1. And I am subtracting this from X 2. So, what now v 2 does not have any component in direction of v 1.

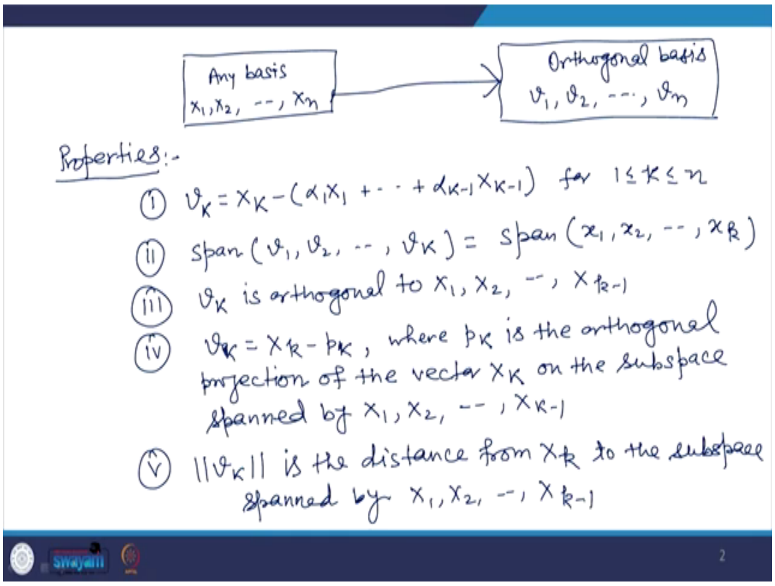
That is component in the direction of v 1 is 0. Hence v 2 v and v 1 are perpendicular that you can see here also. If you find out v 1, v 2 that will become v 1 is your X 1 and this is X 2 minus X 2, X 1 upon X 1, X 1 into X 1. So, this will become X 1, X 2 by this minus X 1, X 1 will be canceled out X 2, X 1 and this is 0. So, inner product of v 1 and v 2 is 0, so, they are orthogonal.

Similarly, I will go for third vector v 3. So, v 3 will be again I will subtract from X 3 the component of X 3 in the direction of v 1 and v 2. So, it will become X 3, v 1 upon v 1, v 1 into v 1 minus X 3, v 2 upon v 2, v 2 into v 2. And I will go in the same way, X 4 like that and finally, for v n I will be having X n minus X n, v 1 upon v 1, v 1 multiplied with vector v 1 minus X n, v 2 and so on.

And finally, I will be having X n, v n minus 1 v n minus 1 v n minus 1 into v n minus 1. And as you have seen here v 1, v 2 are orthogonal, similarly v 3 will be orthogonal to v 1 and v 2 and so on. Then what I can write? v 1, v 2, v n is an orthogonal set. So, this process of finding another set which is orthogonal from a given linearly independent set is called Gram Schmidt process.

Now, what is important here? Important is since X 1, X 2, X n is a basis of V, so, they span the vector space V, the question is whether this new set v 1, v 2, v n is also spanning the vector space V. Yes, the span of X 1, X 2, X n equals to span of v 1, v 2, v n. We can easily prove it.

(Refer Slide Time: 12:32)



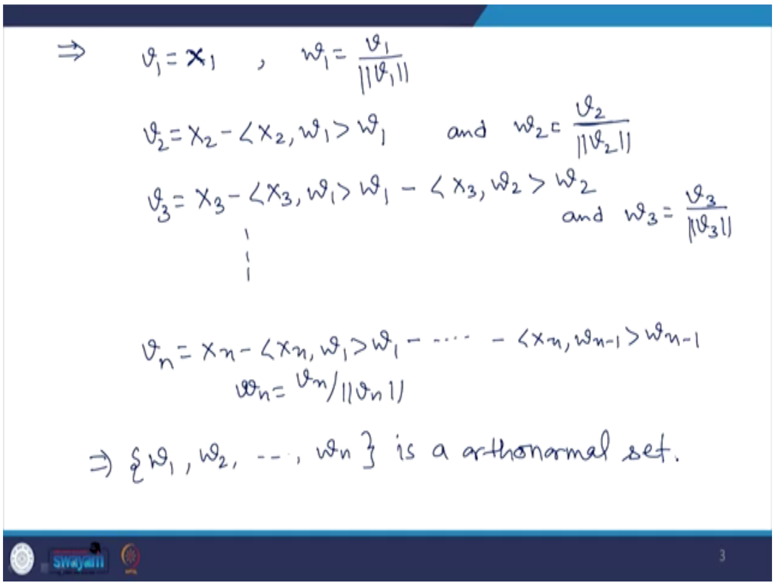
So, what we have seen? Any basis X 1, X 2, X n we apply the Gram Schmidt process on it. And then we got the orthogonal basis v 1, v 2, v n, where v 1 equals to X 1, v 2 equals to X 2 minus the component of X 2 in the direction of v 1.

Similarly, v 3 will be X 3 minus the component of X 3 in direction of v 1 minus the component of X 3 in direction of v 2 and so on, like we have seen it on the previous slide. Now, some properties. So, first property is here v K will be X K minus alpha 1 X 1 plus alpha K minus 1 X K minus 1 for 1 to K to n. As you and what is alpha 1? Alpha 1 will be inner product of X K with v 1 upon v 1, v 1; inner product of v 1 with itself and so on. So, these are the scalar coefficients.

Now, second is a span of v 1, v 2, v K equals to span of x 1, x 2, x k. And this you can prove by this one easily by using this particular thing. Now, third one is v K is orthogonal to X 1, X 2 up to X k minus 1; number 4 v K equals to X k minus p K. So, what I can write? Where p K is the orthogonal projection of the vector X K on the subspace spanned by X 1, X 2, X K minus 1.

And what is this? The length of v K that is norm of v K is the distance and what distance? Shortest distance, from X k to the subspace spanned by X 1, X 2, X k minus 1. So, we can make in this way. So, these are some properties very easy to remember and easily verify from the Gram Schmidt process.

(Refer Slide Time: 17:18)



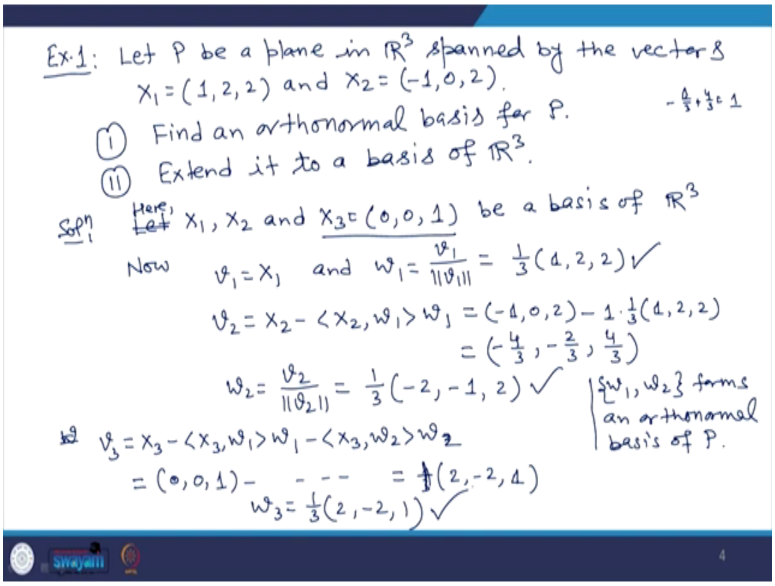
Now, if you write like this just the same process. What we were doing? We are writing v 1 equals to X 1 and, then what you do? You write w 1 equals to v 1 upon norm of v 1. Then what you are doing? You are calculating v 2.

So, v 2 you write X 2 minus X 2 with w 1 into w 1. So, what, the same thing because w 1 is having unit length. So, I am not writing denominator here. So, same thing which we have done earlier only thing in terms of unit vectors. And then what you do? Define w 2 equals to v 2 upon norm of v 2. So, w 2 again a unit vector then v 3 will become X 3 minus X 3 w 1 multiplied with w 1 minus inner product of X 3 with w 2 multiplied by with w 2.

And then w 3 you define v 3 upon norm of v 3. If you go in this way finally, you will be having v n that is X n X n, w 1 multiplied with the inner product of X n and w 1 with w 1 minus and finally, inner product of X n with w n minus 1 multiplied this scalar with w n minus 1. And, then define w n equals to v n upon norm of v n.

Then what you will be having? This side w 1, w 2, w n span the same set which is spanned by v 1, v 2, v n or X 1, X 2, X n and it is a orthonormal set. So, in previous slide we have seen about orthogonalization process using Gram Schmidt, but here we have just modified it in case when you need orthonormal set. So, now we will see some of the examples how to implement this process for finding the orthogonal or orthonormal basis.

(Refer Slide Time: 20:09)



So, let us do example 1. So, let P be a plane in R 3 spanned by the vectors X 1 equals to 1, 2, 2 and X 2 equals to minus 1, 0, 2.

So, now first find an orthonormal basis for the subspace P and the second thing extend it to a basis of R 3. So, we can solve this problem using 2 way; 1, first we will solve this one that we will find an orthonormal basis for P using the Gram Schmidt process on X 1 and X 2. And, then we take a vector orthonormal to w and w 2 which is coming using after applying the Gram Schmidt process.

And, then it will become the basis of R 3 or you first take a li vector to X 1 and X 2 let us say X 3 and then apply the Gram Schmidt process. So, let us solve it. So, let X 1, X 2 and X 3 which is you just take li to X 1 and X 2. So, I am taking 0, 0, 1 be a basis of R 3. So, I have added extend the basis of which X 1 and X 2 of P to a basis of R 3.

Now, so in fact it is not let here, now v 1 will be your X 1. And then w 1 will become v 1. So, what is norm of v 1? It will become 1 9 square root 9 will be 1 by 3 so, 1 by 3 1, 2, 2. Then I will be having v 2 that is your X 2 minus X 2 inner product of X 2 with w 1, multiplied with w 1.

So, here inner product is usual dot product in R 3. So, X 2 is minus 1, 0, 2 minus dot product of X 2 with w 1; so, minus 1 into 1 by 3. So, minus 1 by 3 plus 4 by 3; so, minus 1 by 3 plus 4 by 3 it will become 1. So, minus 1 times w 1 and w 1 is 1 by 3 1, 2, 2. So, minus 1 minus 1 by 3; so, minus 4 by 3 0 minus 2 by 3 and then 2 minus 2 by 3, so, 4 by 3.

Now, calculate w 2; w 2 will become v 2 upon norm of v 2. So, this will become 16 by 9 plus 4 by 9 plus 16 by 9. So, 36 by 9; so, 4 and 4 square root will become 2. So, it will become 1 by 3 minus 2 because I have to divide by 2 minus 1 and 2. So, this is your w 1 w 2. And now w 1, w 2 forms an orthonormal basis of P.

Now, what is orthonormal basis of R 3? So, I will calculate v 3; so, v 3 will become X 3 minus X 3 inner product with w 1 multiplied with w 1 minus inner product of X 3 with w 2 and that scalar multiplied with w 2.

So, X 3 is here 0, 0, 1 minus inner product of X 3 with w 1 and so on and after calculating it, I will get 1 by 3 2, minus 2, 1; so, that will be w 3. So, whatever v 3 I will get means, I have to divide means it will be 2 minus 2 1 and then w 3 will become length of this that is 9 root 9; so, 1 by 3 2, minus 2 and 1. So, here this w 1, w 2 and w 3 forms an orthonormal basis of R 3.

So, in that way we can apply the Gram Schmidt process for finding the orthonormal basis for any subspace of a given vector space, or we can apply it for finding the orthonormal or orthogonal basis for a complete vector space; means. vector space of any dimension. That is finite dimension, but not for in finite dimension.

So, in this lecture we have learned how to solve or how to form an orthogonal or orthonormal basis for a vector space using the Gamma Schmidt process. In the next lecture we will learn another important decomposition of matrices that is polar decomposition. In that decomposition what we usually have? We used to have a complex number like, polar form of a complex number like a decomposition of a given matrix, in form of an isometry and a positive definite matrix, so.

Thank you very much.