NPTEL

NPTEL ONLINE CERTIFICATION COURSE

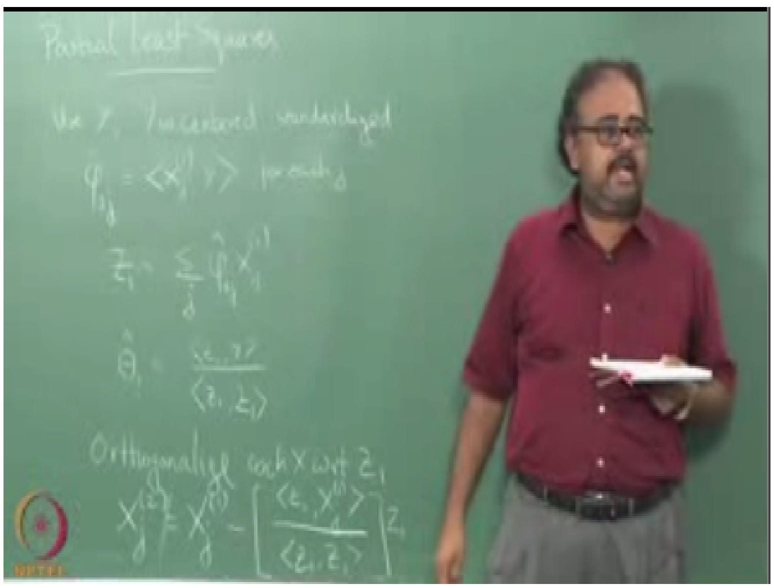
Introduction to Machine Learning

Lecture 18

Prof. Balaraman Ravindran
Computer Science and Engineering
Indian Institute of Technology Madras

Partial Least Squares

(Refer Slide Time: 00:17)



Okay so we will continue from where we left off as I promised right so we are looking at linear regression and we looked at subset selection and then we looked at the shrinkage methods and then, finally we came to derive directions all right I said there are three classes of methods so we are looking at a couple of examples of each of those classes of methods the first one we looked at was subset selection so we looked at forward backward selection and stage way selection in step wise selection and all that and then we looked at shrinkage methods where we looked at rigid regression and lasso and then we started looking at derive directions right.

Where we looked at principal component regression I said the next one we look at is partial least squares and it gave me the motivation for looking at partial least squares it is mainly because principal component regression only looks at the input data okay, does not pay attention to the output right and therefore you might sometimes come up with really counterintuitive directions like an example I showed you with the plus and -1 outputs okay, so the basic idea here is that we are going to use the Y also right.

Just like the usual case I am going to assume that Y is centered right, and I am also going to assume that the inputs are standardized this is something which you have to do for both PCA and partial least squares essentially assume that the each column right it is going to have 0 mean unit variance right on the data that is given to you make it 0 mean unit variance, so that you are not having any magnitude related effects on the on the output okay, so what I am going to do is the following if you remember how we did orthogonalization earlier something very similar so I am going to look at.

So I am going to look at the projection of Y on XJ right then I am going to create a derive direction which essentially sums up all of these projections right I have contributed basically I am computing the projection of Y on xj right, so this is essentially the direction is a vectorized version of it then I am going to sum all of this up so essentially what I am doing here is I am looking at each variable in turn I take each x_j in turn okay I am seeing what is the effect on Y right, so how much of Y I am able to explain just by taking xj alone and I am using all of that I am combining that and making that as my single direction so individually taking each one of this all by itself okay.

Individually taking each direction by itself how much of Y can I explain and that becomes my first the rave direction that is my z1 okay, so that is the coefficient for z1 in my regression for eventual regression fit okay that is the coefficient for that one you can see what it is like so I have taken Y and regressed it on z1 and that essentially gives me what the coefficient for z1 right so how do I go on to fine okay so I am looking at how much of Y is along each direction xj right so in some sense you can think of it as if I have one variable xj right.

How much of Y connects explained with that one variable xj okay I am looking at that and then my first Direction z1 is essentially summing that univariate contributions over all my input directions I suppose, I have to input directions fortunately I have to do this in suppose I have to input directions so what I am going to do is I am going to take my Y right, so project it on x1 alone first right project it on x1 alone and on x2 alone right we do that this is tricky to do this in 3d but any way right.

No it is going to be hard to do it on the board pictorially for you okay I am not going to do this so I really need to get a actually have to plot a function why right I cannot just do it with single data points why that does not make sense, so I actually have to get to a surface Y on x x1, x2 and then talk about the projection so that is going to be hard right, but the basic idea is I take Y right I find the projection of Y along x1 okay then I find the projection of way along x2 okay now I am going to take the sum of these two okay and whatever is the resulting direction and I am going to use that as my first direction.

Yes we see in PCR what we did was we first found directions X which had the highest variance here we are not finding directions in X with the highest variance but we are finding directions in X right some sensor components of X which are have more in the direction of the output variable Y right, so eventually you can show that which you are not going to do but you can show that the directions you pick that Z 1 Z2 Z 3 that you pick or those which have high variance in the input space.

But also have a high correlation with way right it is actually an objective function which tries to balance correlation with Y and variance in the input space well PCA that is only variance PCR there is only variance in the input space does not worry about the correlation but partial least squares you can show that it actually worries about the correlation as well right. We find the first coordinate now what do you do you orthoganalize, so what should I do now I should regress x1 so what should I be doing now I should regress x1 like xj on z1 right.

This is how we did the orthogonalization earlier right, so you find one direction okay then you regress everything else on that direction then subtract from it that gives you the orthogonal direction right, so essentially that is what you are doing here the expressions look big but the expressions look weak but then if you have been following the material from the previous classes then it is essentially whether they just reusing the univariate regression construction we had earlier right.

So now I have a new set of directions which I call xj to write x j1 was the original xj is I start off with now I have a new set of directions which we will call xj2 and then I can keep repeating the whole process, I can take white projected along xj2 to write and then combine that and get Z2 and then regress Y on Z2 to get θ 2 right, so I can keep doing this until I get as many directions as I want all right so what is the nice thing about there is that one get two and other things they themselves will be orthogonal because they are being constructed by individual vectors which are orthogonal with respect to their all the previous rates that we have right.

Each one will be orthogonal and therefore I can essentially do univariate regression so I do not have to worry about accommodating the previous variable, so when it when I want to fit the set K I can just do a univariate regression of Y on that K and I will get the coordinates theta K okay is it fine great. So once I get this theta one to θ K how do I use it for prediction can I just do like xβ like into xβ can I do xθ and I know what should I do well so I can do zx z read it mean I am sorry θ is dead right I can do θ Z and predicted.

But then I do not really want to construct this Z directions for every vector that I am going to get so I do not want to project it along those Z direction, so instead of that what I can do if you think about it each of those threads is actually composed of the original variables X right. So I can compute the θ and then I can just go back and derive coefficients for the XS directly because all of these are linear computations I all I need to do is essentially figure out how I am going to stack all the t-test so that I can derive the coefficients for the XS okay think about it you can do it as a short exercise but I can eventually come up and write right.

So where I can derive this coefficients β hat from these θ right so I will derive data 1 θ 2theta 3 so on so forth I can just go back and do this computation so you will have to think about it very easy you can work it out and figure out what is the number should be right and what is am doing that number of their directions, I actually direct write the number of directions I derive from the PLS right so here the first direction I can keep going suppose I derive P directions what can you tell me about the fit for the data if I get P PLS directions it essentially means that I will get as good a fantastic.

Original least squares fit right so I essentially get the same free as least squares fit okay so anything lesser than that is going to give me something different from the least squares fit okay here is a thought question if my X are originally, orthogonal to begin with next were actually orthogonal to begin with what will happen with PLS will be the same as X is right and what will happen to Z2 can I do the Z2 no right PLS will stop after one step because there will be no residuals after that right so I will essentially get my least squares fit in the first attempt itself okay so that is essentially what will happen right so we will stop with regression methods.

IIT Madras Production

Funded by
Department of Higher Education
Ministry of Human Resource Development
Government of India

www.nptel.ac.in