Modern Computer Vision

Prof. A.N. Rajagopalan

Department of Electrical Engineering

IIT Madras

Lecture-84

Okay, GMM for actually a segmentation. Now so it is like you know a mixture model right. So it is like saying that you know if I, so right instead of you know going, so it is one way to think about it is it is kind of a probabilistic version of actually k-means. Version of k-means. You can also think of it as some kind of you know a generative model okay at the end of the day. But anyway right as far as segmentation is concerned right we can actually think about it as being a probabilistic version of k-means.

And one of the things that it does is clusters are modeled as actually a Gaussian, modeled as a Gaussians which basically means that means that we do not simply take the means alone right. Now we are going to talk about means, covariances everything. And the other thing is that it does a soft assignment right. So in that sense right even if you have something on the boundary and all right you do not have to worry, you can say that there is a higher probability that it belongs to that cluster probably lesser probability that it actually belongs to the other cluster and so on right.

So it does enable soft assignment and then if you actually go through the steps right you will realize that you know that basically what you call a probabilistic thing right will become very very apparent when you compare it with how k-means work. And the cluster shapes are of course right, what will be the shape here? No, no but what will be the shape of the cluster, there it was a center here it will be, there it was spherical right sorry not center what is it ellipsoidal, ellipsoidal. Again right I mean you know again right it does not mean that mean that right GMM can solve for example right all your issues for example if I showed you one such image right which I think if I showed okay anyway right we would not get into that but I am just saying that this is at least better than the k-means because this is more like you know it is taking more things into account right while it is actually doing the assignment and also the cluster shapes right need not be need not be constrained to be spherical which is actually a good thing. So for example right in the other case right where we had the data no where is that we are here right so now I mean I can probably think of right having some kind of an ellipsoidal for this and then maybe write something like that and then I can have a soft assignment for these pixels that are falling in the boundary and I can probably say that with more confidence that they belong to the inner cluster and probably with less confidence that they belong to the outer cluster, this it has to figure out right we do not know because again right it is again the generative in the sense that right there is there is something right something underlying that happened which simply throughout all these points now and right and basically we do not know from which of these Gaussians right these points came right

and our job is to kind of figure out as to as to what is the what is the group assignment and then right in the process arrive at a segmentation task. So yeah so that is indeed having this ellipsoidal of course you know is surely more helpful than sort of you know than a spherical cluster.

Now the way you do is so the way this works is the data sort of a distribution right that from where right all of all of this happens to happens to come all these data points have been observed right is actually described by this one mixture model okay by a GMM which means I mean I am going to take all scalars okay but all of this is easily extendable to the vector case just for simplicity. So P of x right will look like summation over the over let us say c number of the clusters but again right you need to you need to know you need to know these know the number of Gaussians just as in k-means right you need to know the number of groups again here right you need to know the know the know the number. So πc then let us say Gaussian x semicolon whatever μ c σ c and this πc sort of right indicates the strength of the Gaussian of the cth Gaussian right of the cth Gaussian. So again right again again it is the same issue and so for example we want to be able to tell that so we want to be able to model this using μ ltiple such Gaussians and we need to know right what should be the what should be the number of such such Gaussians that we need to model the Cp of x and all that we have is again the same set of points right. Now the way to way to way to write you know walk around this problem of problem of not knowing a group assignment see for example right if you knew the group assignment then maybe you can just do an MLE and then you are done right you will get your you will get your means variances everything right which will optimize for that for that observation but in this case because we do not know right so the way right typically okay it is done is done is that you introduce what is called a what is called a latent variable okay so there is called latent or say hidden.

So what is what is this thing right when which is latent this is πC so for example right so this πC for those of you who have done how do you interpret π C I have just written it as a strength of the Gaussian but then π C is like what I mean I know what would be exactly so it is like the fraction of the point which is in that group right as compared to the entire set of points right. So it actually so in a sense right so π C is π C is very high for I say certain group that it means that when you draw a sample it is very likely that right it will come from there because of the fact that πC for that is very high because among the total set of points that particular group happens to have a lot of points right within it but this right since we do not know and we have to start somewhere right so when you start somewhere so what we do is you know so the way to sort of right think about it is first of all among these mixture components right I would like to first of all identify one mixture right. So, for example, so what you will do is you know introduce a latent or hidden variable Z right where you see a probability that is Z equal to C right is equal to actually this one the πC right so what this means is that probability that you will pick up the Cth mixture right that is the this whole thing is a kind of mixture of Gaussians right so the probability that you will pick the Cth mixture itself is actually πC that means if πC is very high for a certain C then you see probability that you would pick that that particular mixture to draw the sample right is very high and once you once you have once

you have done that right then this then this then this X right given Z becomes simply N of N of right X right μ K I will use C there now right given that given that right now I have said C so it is like you know μ C σ C right. So in a way so this but this but this group assignment and the optimization of the of the model right has to be done in a sort of right you know what you call right iterative way and we call this hidden information because this information we do not have right we do not know how this came out right we do not know from where X came whether a particular X came from πor πwhatever the first cluster or the second Gaussian we do not know right so you sort of right making an attempt to get there and what is hidden right we are trying to say that let us get a reveal it but then right initially what you reveal right may not be true. So in the sense so this is where the EM algorithm comes there is something called expectation maximization and this EM algorithm exactly works on this kind of you know principle that in the in the in the E step right you actually reveal the you know hidden information even though it you know you know it may not be accurate but then you start with something where you where you reveal both both is X and Z and then once you have revealed X and Z then you actually find out find out what are called responsibilities you guys have done right GMM therefore you can so I will just write give you an outline of EM and then once you have the responsibilities for each each sort of say data point what responsibility means that how μ ch a particular cluster is responsible for Xi right.

So you have these data points or responsibilities means how μ ch can you say is a particular cluster or how well a particular cluster explains an Xi right that is the that is the E step and then and then once you once once you have these responsibilities then you can actually go to the go to the M step where you where you can actually do and do a sort of a maxi μ m likelihood where you can actually compute recompute your πC's you can recompute now that you know the you know which cluster is responsible how μ ch for each point you can go back and then compute the πC's the μ C's the you know σ C's and all that and then again kind of say come back and then and then write do this process right iteratively and the proof right we will not go into but then you know but then one can one can show that one can show that at this ultimately maximizes maximizes you see P of P of X for all the points right. That means in a sense that you have identified what P of X will best describe your X1 to whatever Xn right that is that is the idea right. So this is a math is fairly straightforward right so so what you have to do is so let us kind of let us say EM for actually GMM right so this is the expectation maximization there is there is a reason why they call this an expectation at all I would not kind of go into that. So EM right so EM for right GMM I mean EM you can apply in various situations in this case we are applying it for a CGM right and then the E step right which is the estimation step or the expectation step as it is called right so here what you do is the following and now right I will draw the parallel with with k-means okay so now you will be able to appreciate as to why we think of this as a kind of a probabilistic version of k-means okay E step. So in the E step right what you will do is for each data sample for each data sample Xi a drawn as per of course drawn according to see πC right as I said earlier drawn according to πC okay and all this assumes that right you are starting with some initial estimates okay.

So that initial estimate could be from the k-means itself or it could be it could also be random but then if it is random right then there could be there could be local optima issues even k-means right there is no guarantee that right that will be there will be a very good initial estimate but in the absence of anything else you can use k-means to start the initial estimation. So we so find find a responsibility right RIC so what is responsible means that how μ ch is a cluster the Cth cluster responsible for for a data point Xi okay and that that you can you can compute as πC and then write Xi μ C σ C upon summation over let us say C prime πC prime N Xi μ C prime σ C prime right and as you can see summation RIC or over C prime over overall all C primes is 1 right. So okay so in essence so the way to interpret this is that if if let us say if let us say write what do you say so if this RIC right if it actually turns out to be turns out to be high right and then it actually means that means that right that particular sort of a Gaussian right is best able to explain that particular particular sort of you know a data point Xi as compared to the others okay that is what this RIC means it will sum to 1 but then for let us say whichever sort of a cluster it turns out to be high that means that means that cluster is able to actually explain it a lot better than let us say write any of the others okay. So yeah so what this means is that a high a good or high value of RIC okay I will just write this is this is a technical word is actually this responsibilities okay a high value of RIC indicates that that that the cluster C or the Gaussian right in that sense the cluster C best explains Xi compared to others compared to the other Gaussians right okay RIC dash and the end step right once you once you once you have these RICs right so the so the end step right what you can see is that you can actually compute πC okay you know let us first compute something called MC which is the which is the total number of data points not really I mean in a sense okay so we can we can actually compute this as summation RIC right over i. See what this what this really means is see for example right I mean if you kind of you if you relate this to the k means right you can interpret MC as being the being the number of data points right in the in the in the in the you know C at the group right because there RIC would have been 1 or 0 right it is either here or there and therefore when you when you come when you come up with you see here right RIC is like is like is like between 0 and 1 right so that is why it is a kind of probabilistic version whereas there you would have had RIC to be it to be just binary right I mean it will be it will be like a one-shot vector it will be like 0 0 0 somewhere 1 0 0 0 right and so wherever it goes right there you there you get a 1 therefore if you sum up the RICs if you think of this step is exactly what we did in the k means except that there RIC was binary right and then when wherever wherever it was assigned it you would simply add add all those points whether here you add up RICs okay.

So so you can kind of draw a parallel between this and this and the hard assignment this where this where it becomes a soft assignment now whereas there it was a hard assignment and then you can then compute πC right you can actually use this MC to actually compute πC so that will be like MC by M where M is the total number of points right and then you can of course compute your μ C which will be like 1 by MC but now but each of the each of the xi should be weighted by RIC RIC xi over i and σ C will also or σ square C will also be like 1 by MC then summation weighted by RIC and then xi minus μ C square right. Now so so so these two steps right if you actually write iteratively do then there is a proof that that this actually

maximizes so so write iteratively doing this right so iteratively iteratively doing E and M step okay and and that is the general EM algorithm okay that is not even for this particular case okay and and M step right will actually will will eventually maximize maximize log likelihood likelihood of maximize the p of x let us say just to be simple okay. So so in a sense you can look upon this as a kind of a generative model right that means that means you got you computed the Gaussians that can actually that can actually best explain right what is happening there okay so yeah right so this in a sense is actually GMM right and then you can show that right if you have this you can actually do a lot better than and as you can see everything is a soft assignment now right through the RIC. So so we are not saying that take the right something is only here or it is only in that group right we are saying that we are saying that right that that there is an there is there is a probability and therefore right you could also have situations like these where you can say that right even points that fall in the boundary right depending upon how this whole thing worked out right you might be able to say with with less confidence that it is in a particular cluster and with more confidence that it probably should belong to the other cluster and so on. And and this choice of feature space right as I said right it is it is an important thing right we will we can just go through the slides this seems to just says that right what if you chose color what if you chose just a gray scale what if you chose the location also along with the intensity this is simple right simply means that you know that feature vector that you have whether you want it to be just a gray scale whether you want it to be RGB whether you want it to be RGB x, y right whether you want to bring in the spatial location into account all of that is up to you okay it depends upon which problem right you are trying to solve okay and and for example if you had something like this right then of course you would have to not just using color and all right would not actually make sense right so you will have to go for a texture or something.

So all that is there right you will have to look at say texture similarity similarly right here a zebra right you will have to use a texture in order to say what it is rather than use something like this which is a color bayer which does not seem to tell anything at all right so therefore which feature space you choose is one thing okay that is a problem in itself and and for that right there is actually enough enough information available okay one can kind of go through any of these works and that will tell you what is a good feature to use but after that right which one should you be using if you go for k-means then certain things will happen if you go with the GMM right then certain things will happen so if you take a GMM right there is basically that is how it will try to model you have these points right it will try to you know fit the ellipsoids through them and yeah this is what it is but I do not I do not know whether there are enough examples here okay. Now the last thing right within this segmentation right that is kind of clustering is what is called mean shift see for example so if you had something like this right and and for example right if you had to do do a clustering this is this is this is this is very hard right for example you cannot think of a GMM being able to do this you cannot think of think of k-means or GMM or anything right that can actually do this and and first of all right how do how do you even tell what is a k and all so so mean shift is actually is actually a higher level algorithm than both of these that is actually I mean earlier one I mean the GMM is actually a parametric approach whereas this mean shift is actually a non-parametric

approach it does not need the value of k does not need anything does not even assume the shape and nothing it assumes okay and that is why it is actually a powerful technique unfortunately not μ ch is talked about mean shift but actually it is a very good a powerful technique so I thought right I will just talk about it briefly okay tomorrow.