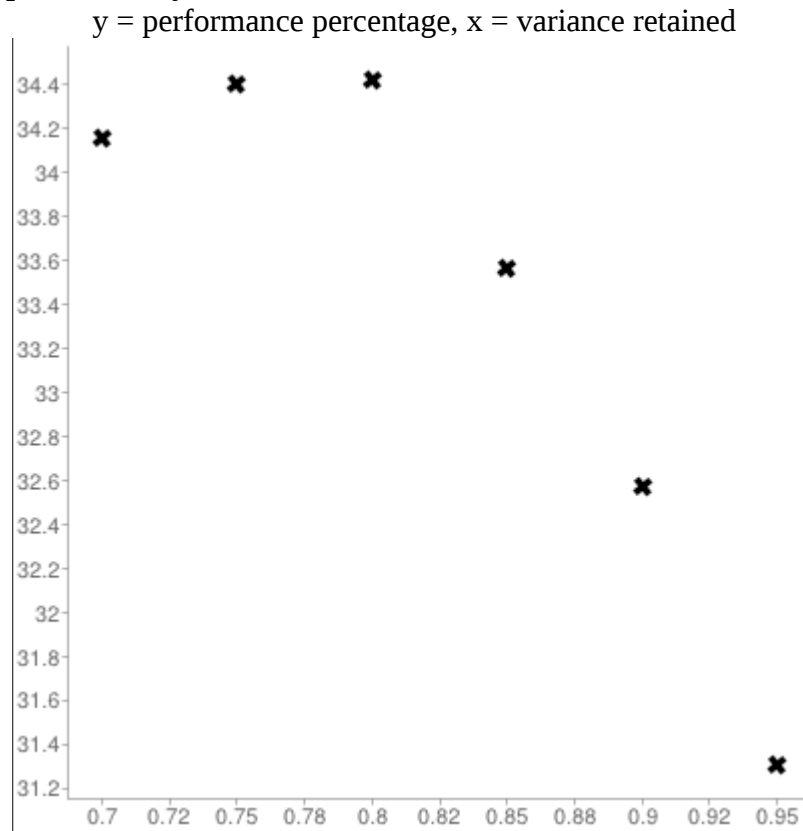


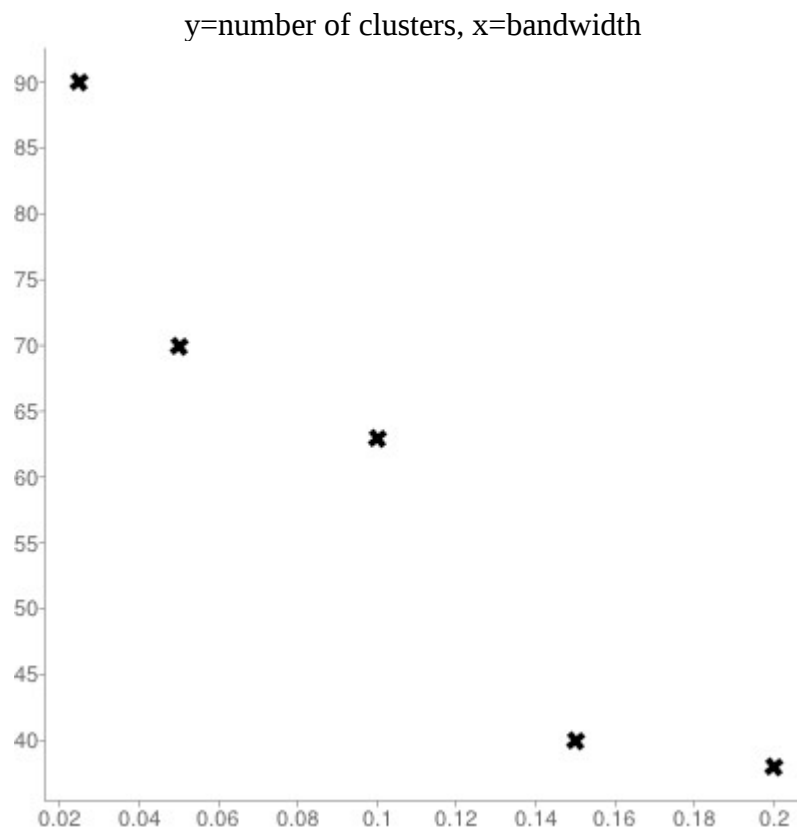
1) Principal Component Analysis



Performing PCA on caltech101 dataset gave better results than histogram intersection and chi-square method. But as you can see in the graph, increasing variance parameter has decreased the performance. I think this is due to irrelevant features in the dataset, as we include more features, it becomes less effective. Because if you examine the categories, even two images from the same category can be totally different images (e.g BACKGROUND_google, stapler). Therefore, more features will be more misleading for PCA in the caltech101 dataset.

I have also provided results.txt in the zip which shows the results of PCA performance.

2) Mean-shifting



Mean-shifting part was pretty challenging. Because there are different parameters to give besides bandwidth. I had to find appropriate thresholds by trial-and-error method. There are two more parameters besides the bandwidth: merge threshold, step threshold. Step threshold decides when to end the iterations, in other words it decides if a point has converged, I set this to 0.01. So in this case if new weighted mean minus previous weighted mean is below 0.01, we decide that the point is converged. I put every converged point into appropriate bins. That is when merge threshold is needed. We can't simply say that every converged point is a cluster itself because lots of them are very close by to each other. In order to merge close points into a cluster, I used merge threshold. I set it to 0.0003, that means if the latest added point is closer than 0.0003 to any point in the bins, don't create a separate cluster for it. As a result I got the graph above. Of course the thresholds still need a fine-tuning. But it was enough to show the idea. As it can be seen, supplying bigger bandwidths decreases the number of clusters. I think it is simply because that a bigger bandwidth includes more points in our dataset, therefore it is more likely to find less local maxima with a wider window. Because all of the points have large weight and they tend to converge at one point in a wider window. If we provide smaller bandwidth parameters, which means smaller window, it is more likely to find more local maxima points as the points will tend to converge at different points which means they will separate.