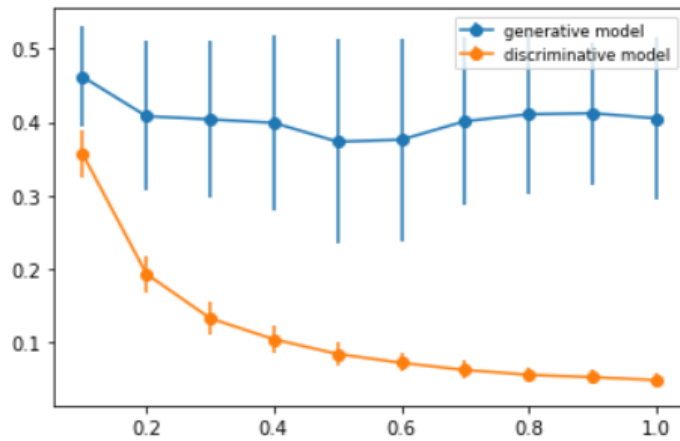


Programming assignment 3

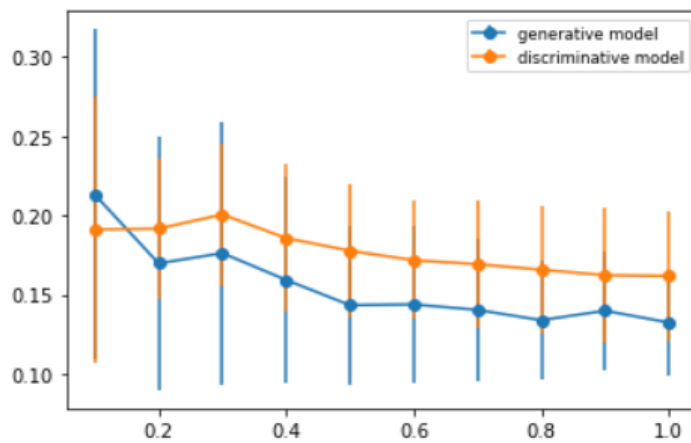
Task1:

In your submission plot these results and discuss them: how do the algorithms perform on these datasets? are there systematic differences? and how do the differences depend on training set size?

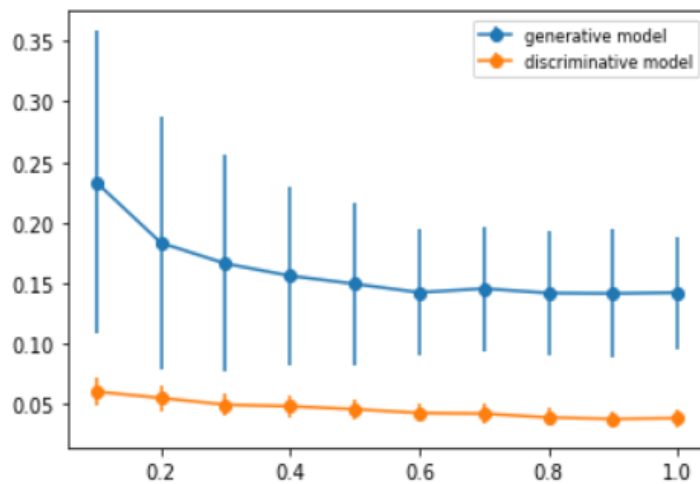
For data set A



For data set B



For data set usps



How algorithms perform on this data set and systematic differences:

For data set A, the Bayesian model and generative model accuracy increase with training size. Also, the standard deviation is very minimum for the Bayesian model compared to the generative model. For both models, as the training size increases, the standard deviation reduces.

For data set B, both models gave similar error rates. the overall error rate is slightly reduced as training size increases. Since the dimension of the B data set is very low compared to the other data set and this dataset generated from multiple Gaussians with differing covariance structure the accuracy is also low, and also the variation of results is very high for both models. But as training size increases, the standard deviation is reduced compared to initial values.

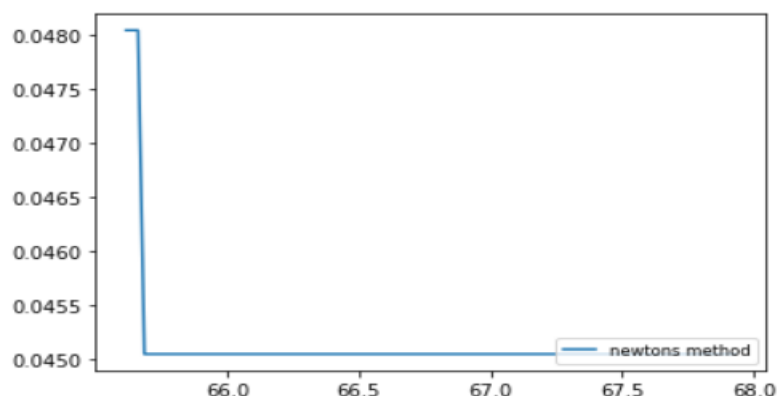
For data set USPS, which is real data, the Bayesian model from the start shows a very low error rate even with less training size, and with training data size increases, the error rate further reduces and accuracy is also very good. In the case of a generative model, the error rate is reduced comparatively but the standard deviation is higher here.

how do the differences depend on the training set size:

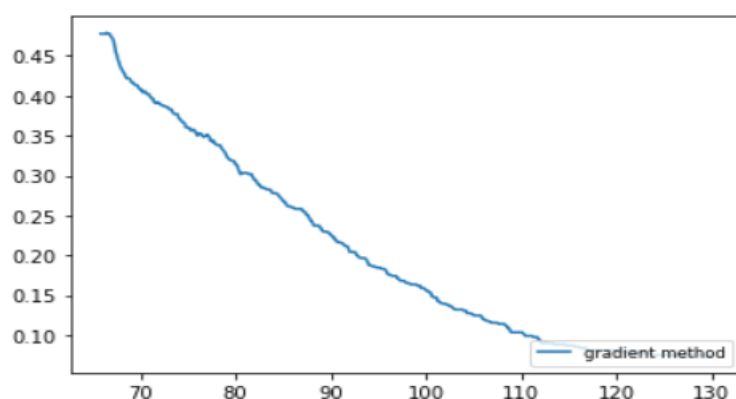
Irrespective of the model/data set, the similarity that is observed is as the training data size increases, the error rate is reduced comparatively. The Bayesian or the discriminative model with a comparatively large number of dimensions data gives a very less error rate even with low training size. Also, the deviation of results from the average value also reduces with training size increases.

Task2:

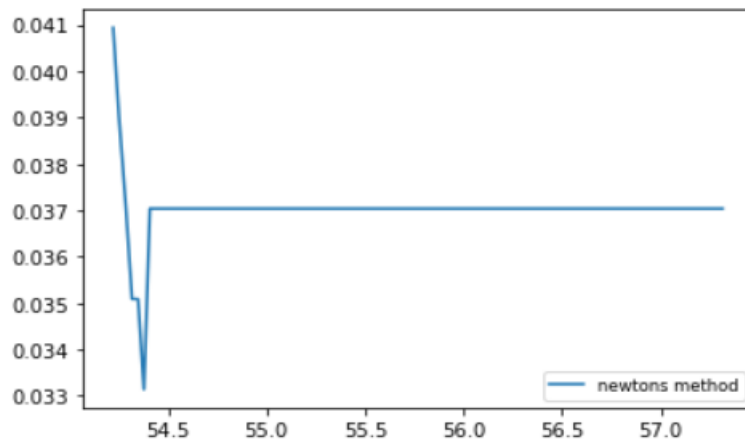
For set A



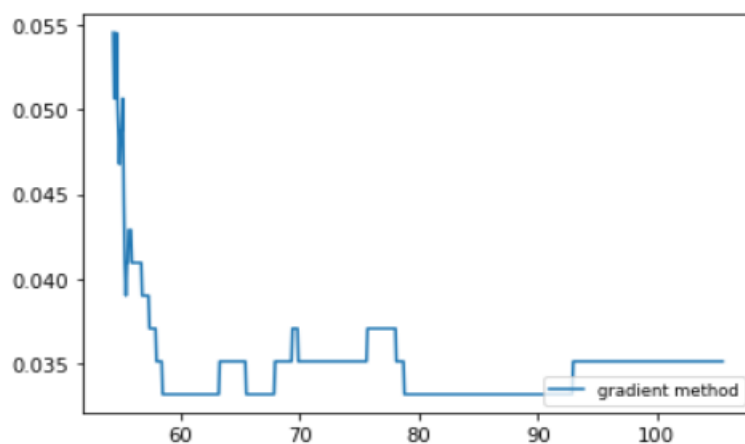
For set A



For set USPS



For set USPS



how do the algorithms perform on these datasets? are there systematic differences? and how do the differences depend on data set characteristics?

how do the algorithms perform on these datasets and their systematic differences:

The total time consumption of the Newton method is less compared to the gradient method even though the complex calculation is not involved in the gradient. This happened because the gradient took so many steps to compute the final output while newtons method computed the same output at very less steps. Also in both cases, the error rate is reduced to very low in the Newton method after very few steps itself, but for the gradient method, the error rate is reduced gradually.

how do the differences depend on data set characteristics:

For the A data set, newtons method converges to a minimum error rate very quickly but for the USPS though it reduced to minimum value very quickly the final error rate is slightly higher than the lowest error rate.

For the A data set, the gradient converges to a min error rate gradually and consistently. In the case of USPS data, as steps proceeds, the error rate slightly increases and then decreases like a zig-zag before converging to a lower value. Here also, the final error rate is slightly higher than the lowest error rate reached by the gradient.

LINE SEARCH:

investigate the notion of line search and then implement and evaluate its effect on the success and convergence speed of gradient ascent.

After initializing the learning rate,

Determining the error rate at every step and if the new error rate of w_{n+1} is higher than the previous one, the learning rate is reduced and this is continued until the final maximized weights are determined.

The code for this is available in the attached pdf and in the .python file.