# Handwritten Digit Classification

## K-NN Classifier:

Sweep K values

for K=1

| Train accuracy | Test accuracy |
|---|---|
| **1.0** | **0.96875** |

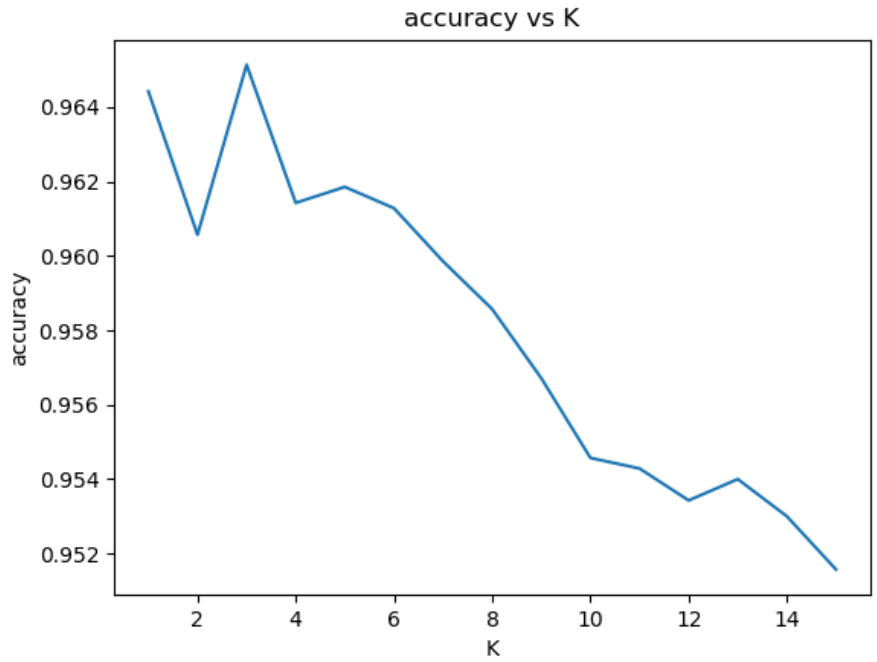
for K=15

| Train accuracy | Test accuracy |
|---|---|
| **0.9658571428571429** | **0.959** |

Strategy to break the tie

In order to break the tie, my method is to decrease the K number by one until there is only one true corresponding class. Since K-NN classifier applies the majority vote scheme, this method guarantees to find only one true class without modifying the existing weighting scheme.

Plot of accuracy vs K for 10 fold cross validation

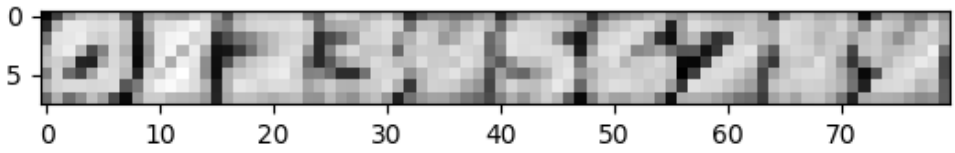



The optimal K is 3 from the plot above.

| Train accuracy | Average accuracy for K-fold | Tet accuracy |
|---|---|---|
| 0.9864285714285714 | 0.96514285714285708 | 0.9705 |

## Conditional Gaussian Classifier Training:

image of the log of the diagonal elements of each covariance matrix




average conditional log-likelihood for train and test set

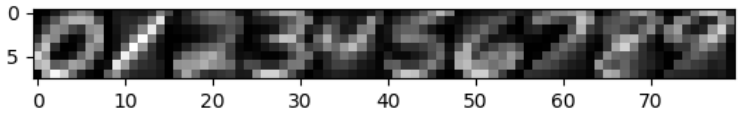
| Train | Test |
|---|---|
| -0.124624436669 | -0.196673203255 |

accuracy for the most likely posterior class for train and test set

| Train | Test |
|---|---|
| 0.9814285714285714 | 0.97275 |

## Naive Bayes Classifier Training:

plot of 8 by 8 grayscale image for eta parameter




sample of one new data point for each of 10 digit classes

Average conditional log-likelihood for train and test set

| Train | Test |
|---|---|
| -0.9437538618 | -0.987270433725 |

accuracy for the most likely posterior class for train and test set

| Train | Test |
|---|---|
| 0.7741428571428571 | 0.76425 |

## Model Comparison:

| Model | Performance |
|---|---|
| K-NN Classifier | K-NN can achieve high classification accuracy (~96%) but there are some issues in this model. First, since there is no training process performed, a very hard database is required to store every training data. Moreover, due to the curse of dimensionality, the computational cost is very high if there are multi-dimensional features. The runtime for each prediction is relatively longer than other classification models. Also, K-NN is not a mathematical model, building a subspace from K nearest neighbors is not an adequate way to represent the whole training set. |
| Conditional Gaussian Classifier | Conditional Gaussian Classifier achieves the higher classification accuracy (~97%) than K-NN Classifier. CGC requires more complex calculations for the covariance matrix but it reduces the overall classification runtime with pre-trained parameters. Also, by applying the Gaussian distribution, the classification decision boundary is non-linear, an advantage to build a more robust multi-class classification model. |
| Naïve Bayes Classifier | Naïve Bayes Classifier achieves the lowest classification accuracy (~77%) among the 3 models. This method assumes the conditional independence for all the pixel points; thus generate the least complex model. However, this IID assumption is generally not true as each pixel is related to the surroundings in a given class. Furthermore, by applying the Bernoulli distribution, the pixel intensity values are being ignored, and this can lead the poor classification performance. |

To sum up, Conditional Gaussian Classifier performs the best while Naïve Bayes Classifier performs the worst. This result matches my expectation as CGC provides a more complete/complex mathematical model and NBC ignores training details with IID assumption.