Introduction

Data

- Exploratory analysis
- Useless variables
- Majority class prediction

Experimental Analysis

Classification with different features

Ink feature

- Explanation of the feature
- Explanation of mean and std + table with values
- Mean values
  #Pairs with the big difference between their mean values can be classified better because it means that in total one digit has more ink than the other.
  [0,1] [8,1] [2,1]
  In general, everything with 1. We can try classification with only one and zero in order to prove our point. Also, we can add images that show that 0 has more ink than 1.
  #Pairs with similar mean values will be classified harder which make sense because all the digits will have close ink values.
  [4,9,7] [2,8]
- Std values
  #large std means that values are more spreaded out which means that some observations are with more and some with less ink (maybe it's the size of the digit so we can find 2 observations of the same digit that we can see the difference).
  # small std it means they are all almost the same.

Classification using ink

- Mention:

  Since in this part of the assignment we only consider very simple models, you may use the complete data set both for training and evaluation.

  Scaling (however even the teacher doesn't know why)
  Fine tuning using grid search cross validation for C values.
- Using multinomial.
- Results
  Accuracy, precision, recall
  22% Which make sense because most of the digits have the same ink value.
  Confusion matrix (we can write some observations about which pair of digits confuses more)

Center mass feature (also we can add mirroring too if we want)

- Feature explanation and why do we think it might discriminate well between the digits.

Classification using mass

- Mention:

Since in this part of the assignment we only consider very simple models, you may use the complete data set both for training and evaluation.

Scaling (however even the teacher doesn't know why)
Fine tuning using grid search cross validation for C values.

- Results
  Accuracy, precision, recall
  Compare ink and mass results.
  Confusion matrix (we can write some observations about which pair of digits confuses more).

Classification using Combination of features.

- Mention:

  Since in this part of the assignment we only consider very simple models, you may use the complete data set both for training and evaluation.

  Scaling (however even the teacher doesn't know why)
  Fine tuning using grid search cross validation for C values.
- Results
  Accuracy, precision, recall
  Compare ink and mass with the combination results.
  Confusion matrix (we can write some observations about which pair of digits confuses more).

Classification using raw pixel as fatures

Data preprocessing

- Mention:
  Using the raw pixel values as features
  Divide the data int 5k for training and 37k for testing using cross validation (folds number)
  Reduce the size into 14x14 in order to reduce the complexity
  Scaling

Classification using multinomial

- Mention:

  Fine tuning using grid search cross validation for C values.
  Lasso for penalty
- Results
  Accuracy, precision, recall
  Confusion matrix (we can write some observations about which pair of digits confuses more).

Classification using SVM

- Mention:

Fine tuning using …

- Results
  Accuracy, precision, recall
  Confusion matrix (we can write some observations about which pair of digits confuses more).

Multinomial vs svm

- **Mention**

  Which is better and why

  Compare the accuracies with a statistical test

Conclusion