

## A KNN

*2 points*

- How many distances you need to calculate if you have 60,000 samples in the trainingset for 50 samples?
- How many distances do you need to calculate if you have  $n$  samples in the trainingset?

## B KNN MNIST

*3 points*

- What is the error rate of KNN on the test set?
- What is the error rate for each label (number)?



Do for  $k = 2, 4, 8$

- How does the choice of  $k$  influence the results?

## C K-means MNIST

*4 points*

- What is the majority class of each cluster?
- What is the percentage of the majority class in each cluster?
- Does each number have a cluster?
- If not, which hasn't?



Do this for 10, 100, 1000 iterations

## D MNIST understanding

*5 points*

Some numbers are more difficult to predict with KNN and to cluster with K-means. Show why. Start with examples, and get more general



## E KNN and K-means CIFAR

*5 points*

transform cifar-10 to grayscale.



- Does knn work similarly good?
- Does k-means work similarly good?
- Demonstrate this similar to B, C and D

## F Linear regression BOSTON

*5 points*

Train a linear regression model on the BOSTON dataset to predict median values of houses. Test it on the test set. Show the code how you did it! What dependent variables (columns) have the biggest influence?

## G Logistic regression MNIST

*4 points*

Evaluate logistic regression as B and D on MNIST



## H Feedback

*2 points*

- How much time did you work on this assignment and how often did you meet?
- Did you use Google Colab or Anaconda?
- What was your favourite exercise? Why?
- What exercise did you like least? Why?