

## MNIST data

- Vi ska jobba med MNIST Handwritten Digits som dataset! Varje siffra består av 784 pixlar och har labels 0-9. Det finns 42.000 datapunkter, dvs handskrivna siffror
- Sätt upp en kNN-classifier för detta och mät score för lite olika värden på k (det vill säga antalet grannar man använder för klassificering)
- sklearn har load\_digits() som snabbt verktyg för att få in datan. Annars finns det på Kaggle <https://www.kaggle.com/c/digit-recognizer/data>
- Snabbrepitition för kNN - <https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761>
- Hinner ni långt här kan ni även göra en "grid search" för att testa scores på olika värden av k. Det ni gör då är helt enkelt att sätta upp olika värden för k i en lista eller annan iterable och mäta score för varje k.
- Vill ni testa en annan algoritm istället kan ni t.ex. använda SVM (Support Vector Machines) som också finns tillgängligt i sklearn. Den borde svara bra på dimensionsreduceringen vi kommer att göra senare.
- Efter PCA – ta och mät tiden det tar för kNN att göra klassifications för olika värden på k och n\_components.
- Gör en confusion matrix för att se var klassificeringarna hamnar och var
- Lägg till confusion matrix till predictions
- Lägg även till så att ni kan variera mängden inputdata
- Lägg till så att tiden sparas för varje inställning med hyperparametrar (n\_components, antal neighbours) samt olika storlekar på inputdatan

