Practical Machine Learning with Functional Programming

NDC mini, 14-15 mars 2018

Intro

- “Functional Programming is a huge topic”
- “Machine Learning is a huge topic”
- “Two days isn’t much time...”
- 10 minuter är ännu mindre tid...

Machine learning

- Skriv ett program för att utföra en specifik uppgift
- Med mer information kommer programmet genomföra uppgiften bättre
- …utan att behöva ändra i koden
- På vissa sätt mindre komplext än det kan verka

Kategorisering

- Classification = Välj ett av ett givet antal val
- Regression = Förutspå ett värde

Kategorisering

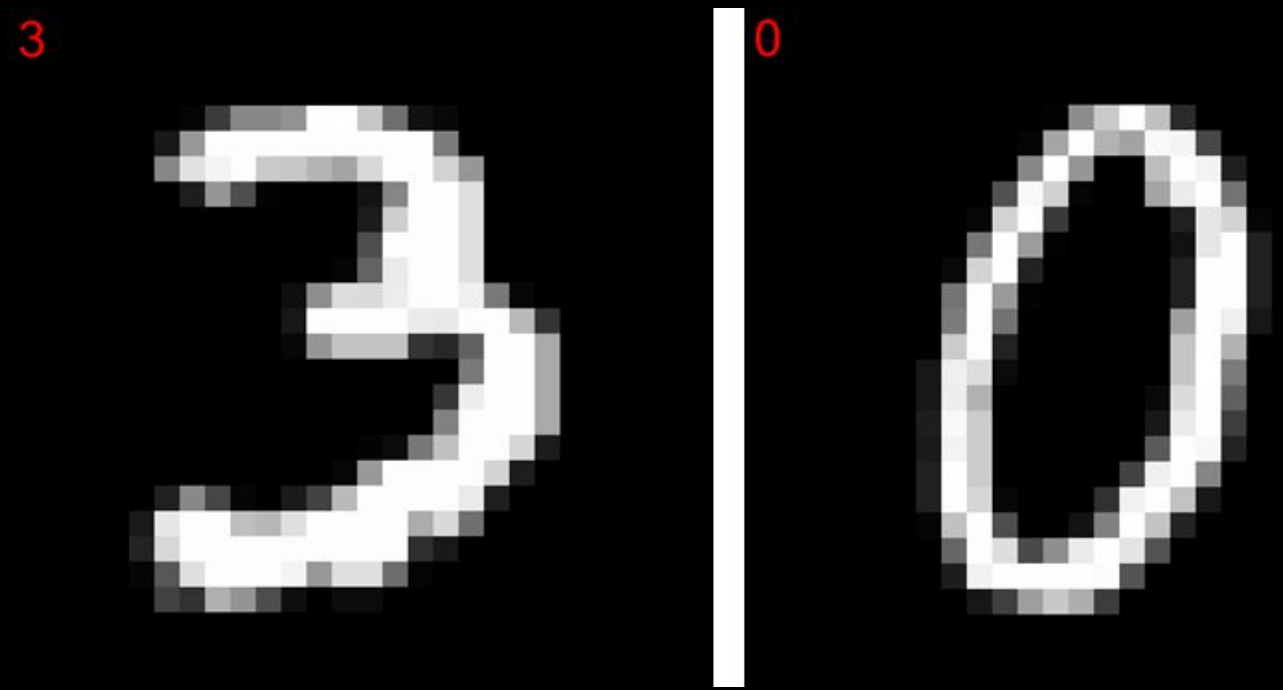
- Supervised learning
- Unsupervised learning
- Online learning

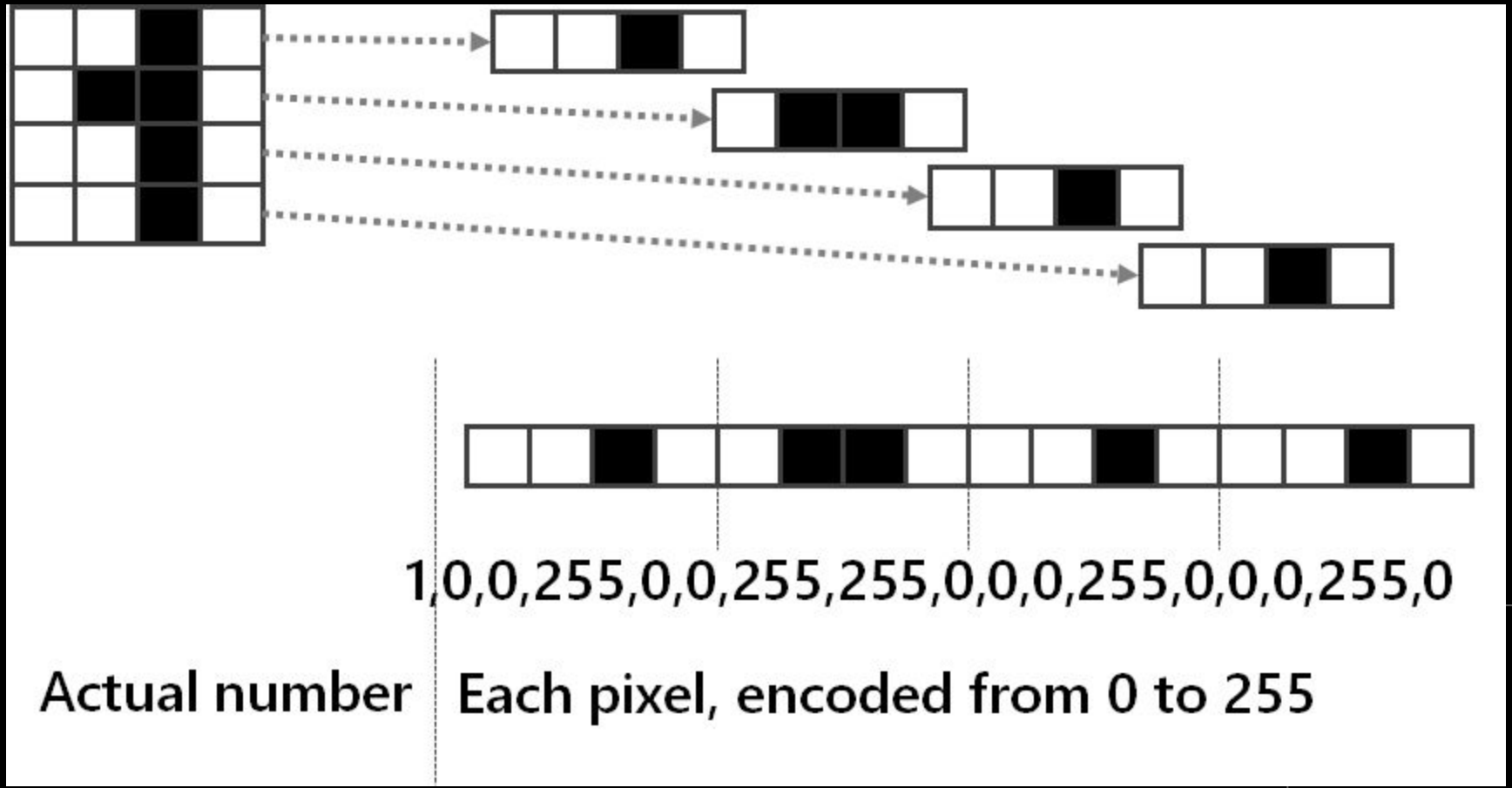
Kaggle digit recognizer

- Dataset med handskrivna siffror
- Mål: Kunna automatiskt identifiera en handskriven siffra
- Exempeldata: 50000 siffror

- Alltså en classifier, dvs:
  - Ge mig en okänd datapunkt så skall jag förutse vilken klass den tillhör
  - I detta exempel har vi klasserna 0-9 (alla siffror)

Datan ser ut på följande sätt




1,0,0,255,0,0,255,255,0,0,0,255,0,0,0,255,0
Actual number
Each pixel, encoded from 0 to 255

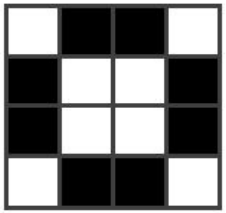
KNN classifier

- **K** **N**earest **N**eighbour
- Givet ett okänt ämne för att klassificera
- Hitta alla kända exempel
- Hitta de **K** närmaste exemplen
- “Majoritets rösta”

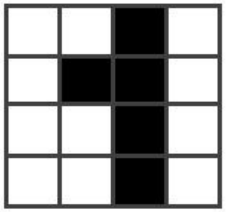
Illustration: 1-nearest neighbour

Sample

0

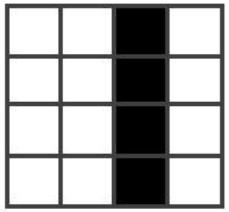


1



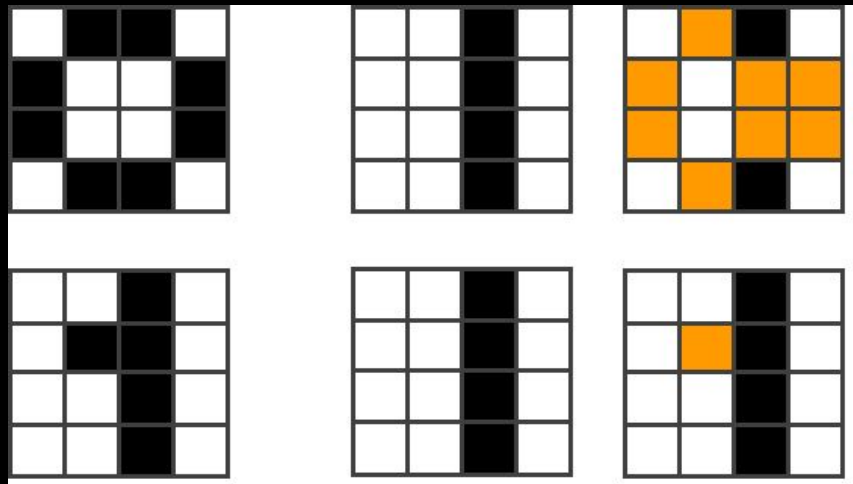
Unknown

?



Which element in the Sample is the most similar / closest to the Unknown item we want to classify?

Illustration: 1-nearest neighbour



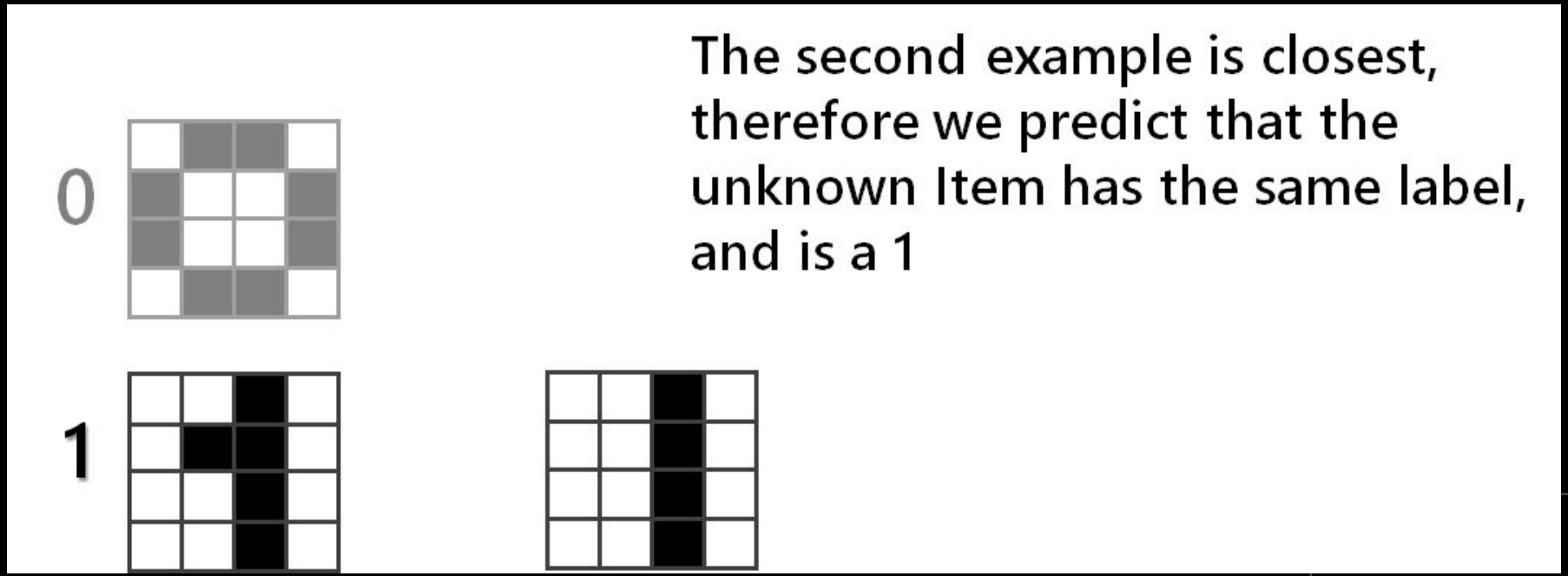
Compare images, pixel by pixel

$$D = \sqrt{255^2 + 255^2 \ldots + 255^2}$$

$$D = \sqrt{255^2}$$

We compute the distance between each element of the Sample, and the Item we try to classify

Illustration: 1-nearest neighbour



Metodik

- Se till att ha data för träning och validering
- Skapa först en enkel och fungerande modell som “baseline”
- Därefter se om den är tillräckligt bra. Kan hastighet förbättras, kan den göras mer precis?
- Hur man beräknar avstånd är avgörande

Övrigt som gicks igenom

- Språkigenkänning
- Beräkna vilket vin som är “bäst” baserat på olika attribut
- Förutse hur många cyklar som används under en dag utifrån specifika förutsättningar