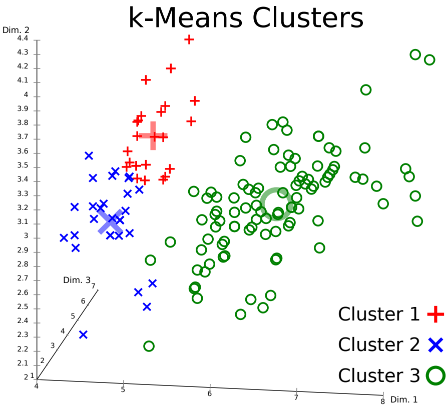
Machine learning

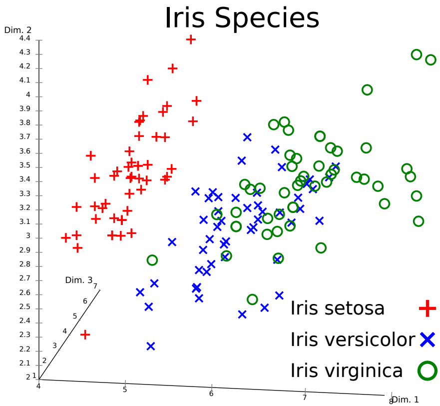
Problem definition

- Formulate learning problem.
- Obtain data.
- Run algorithm on data.
- Obtain conclusion.

Algorithms vary depending on the learning problem.

A Supervised and an Unsupervised Learning Problem





Supervised vs Unsupervised Learning Problems

The clustering problem

- Input: Data $\mathcal{D} = (x_1, \dots, x_N)$, $x_t \in X$, $K \in \mathbb{N}$
- Output: Centers $\bar{x}_c$, labels $(y_1, \dots, y_N)$, $y_t \in \{1, \dots, K\}$.
- Objective (example): minimise intraclass inertia

$$\sum_{c \in [K]} \sum_{t : y_t = c} \|x_t - \bar{x}_c\|^2.$$

The classification problem

- Input: Data $\mathcal{D} = ((x_t, y_t))_{t=1}^N$, $x_t \in X$, $y_t \in Y = \{1, \dots, K\}$
- Output: Classification rule: $f : X \to Y$.
- Objective (example): Minimise classification error

$$\sum_{t \in [N]} \epsilon y_t, f(x_t)$$

Food for thought

Are these the right objectives? What are potential flaws?

Unsupervised learning problems

Problem characterisation

Find a model describing the data.

Example problems / Description

- Clustering / clusters
- Data compression / compressed data
- Density estimation / probability density function
- Document analysis / document topics
- Network modelling / links between entities
- Preference elicitation / user preferences

Supervised learning problems

Problem characterisation

Find a function $f : X \to Y$ making predictions from partial information

Example problems / Functions

- Classification / map from observations to classes
  - Speech recognition
  - Image classification
- Regression / Find $f : \mathbb{R}^n \to \mathbb{R}^k$.
  - Risk analysis
  - System dynamics
- Sequential prediction / map from past to future observations

Training and testing

Measuring objectives

- Say $\bar{x}_i$ are cluster centers minimising an objective for data $\mathcal{D}$
- Do they also minimise it for data $\mathcal{D}'$?

Holdout sets

- Minimise objective on $\mathcal{D}$ and compare with objective on $\mathcal{D}'$.
- $\mathcal{D}, \mathcal{D}'$ can be obtained by splitting the original data in two parts.

Example on intraclass variance for kmeans

- What is the expected behaviour in $\mathcal{D}$ and $\mathcal{D}'$?
- What actually happens? How can we explain it?

The importance of the objective function

Remember that the original objective is

$$\sum_{c\in[K]} \sum_{t:y_t=c} \|x_t - \bar{x}_c\|^2.$$

Let's try and implement an alternative objective

$$\sum_{c\in[K]} \sum_{t:y_t=c} \frac{1}{N_c} \|x_t - \bar{x}_c\|^2.$$

The simplest classifier

$y = \texttt{Look-Up}(x, \mathcal{D})$ // Data $\mathcal{D}$, new point x

for $(x_t, y_t) \in \mathcal{D}$ **do**
  if $x_t = x$ **then**
    return $y = y_t$.
  end if
  return $y \sim \mathcal{U}\mathit{nif}(Y)$.
end for

Definition 1 (The uniform distribution)

If $\mathbb{P}$ is the uniform distribution on Y, then

$$\mathbb{P}(A) \leq \mathbb{P}(B) \Leftrightarrow |A| \leq |B|, \qquad A, B \subseteq Y$$

Sp: For $\mathcal{U}\mathit{nif}(\{1, \ldots, N\})$, we have $\mathbb{P}(k) = 1/N$.

- Identify one or more weaknesses of this classifier.
- How could this classifier be improved?

The simple multinomial classifier

$y = \texttt{Multinomial}(x, \mathcal{D})$ // Data $\mathcal{D}$, new point x

1: **return** $y \sim \mathcal{M}ult(p(x))$,

$$p_i(k) = |\{x_t = k \wedge y_t = i\} / |\{x_t = k\}|$$

The estimate is the proportion of data with $x_t = k$ which have label i.

Definition 2 (Multinomial)

If $y \in \{1, \ldots, K\}$ is multinomially distributed with parameter $p \in [0,1]^K$, $\|p\|_1 = 1$, we write $y \sim \mathcal{M}ult(p)$. The probability that y takes the value i is p_i.

How can we generalise this?

- What about not previously seen x?
- When x is continuous.
- When $x = x(1), \ldots, x(n)$ is a long vector of features.

How can we generalise this?

- What about not previously seen x?
- When x is continuous.
- When $x = x(1), \ldots, x(n)$ is a long vector of features.

Some algorithms

- Decision stumps
- Decision trees
- Nearest neighbours
- Bayesian networks
- Support vector machines