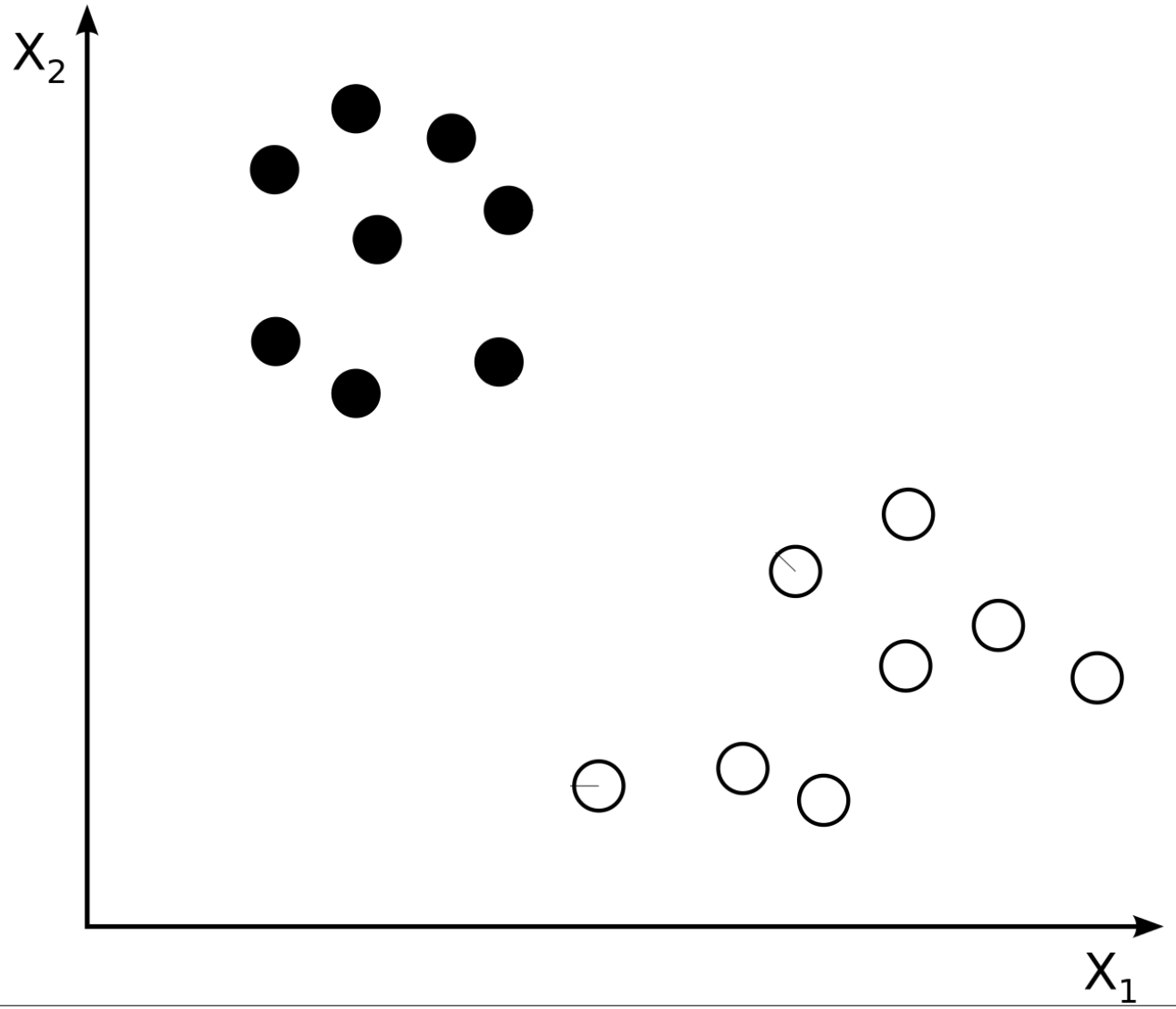
Machine Learning

Support Vector Machines

Support Vector Machine

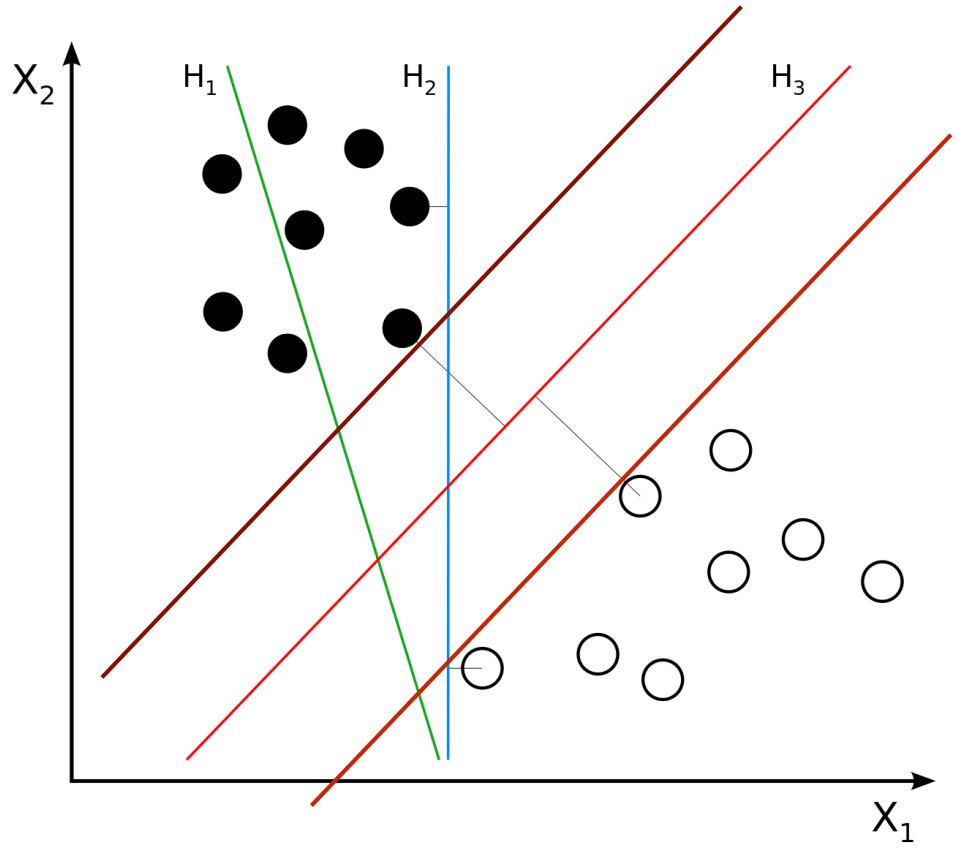
- A Classification Problem
  - Linearly separable



Source: Wikipedia

Separating Hyper-planes

- There are many hyperplanes that might classify the data.
- One reasonable choice as the best hyperplane is the one that represents the largest separation, or margin, between the two classes. So we choose the hyperplane so that the distance from it to the nearest data point on each side is maximized.
- If such a hyperplane exists, it is known as the maximum-margin hyperplane



Source: Wikipedia

Hard Vs Soft Margin

- If the data is linearly separable
  - we can select two parallel hyperplanes that separate the two classes of data, so that the distance between them is as large as possible.
  - The region bounded by these two hyperplanes is called the "margin", and
  - the maximum-margin hyperplane is the hyperplane that lies halfway between them.
- When data are NOT linearly separable
  - Hinge loss function: adds a penalty for crossing over the margin
  - Penalty is proportional to the distance from the margin

Kernel SVM

- “A complex pattern-classification problem, cast in a high-dimensional space nonlinearly, is more likely to be linearly separable than in a low-dimensional space, provided that the space is not densely populated” -- T. Cover

Kernel SVM

- "A complex pattern-classification problem, cast in a high-dimensional space nonlinearly, is more likely to be linearly separable than in a low-dimensional space, provided that the space is not densely populated"
  - Tom Cover