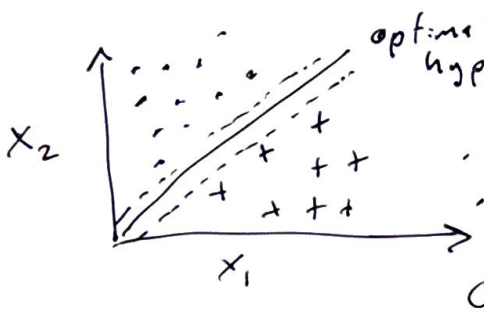
Supervised Learning: Linear and Non-Linear Support Vector Machines



- Linearly seperable classes
- Multiple ways to separate the classes

Optimal hyperplane ~~maximizes~~ maximizes the margin to separate the two groups.
- Exactly at the middle of the two groups

Linear Support Vector Machines

- Popular approach in research, but not in industry
- Simplest case: maximize the margin - the distance between the boundary (hyperplane) and the support vectors (training examples closest to the boundary)
- Max margin picture not applicable in non-seperable case
- scikit-learn: sklearn.svm.SVC

Sometimes the classes can be seperated, but not linearly: Non-Linear Support Vector Machines

- Also popular in research, not industry
- "Kernelize" for non-linear problems
  - Choose a distance function named a "kernel"
  - Map the learning task to a higher dimension
  - Apply a linear classifier (SVM classifier) space in the new space
- Not memory-efficient, because it stores the support vectors, which grow with the size of the training data
- Computation is expensive
- sklearn.svm.SVC