Introduction to Machine Learning Applications

Spring 2023

Support Vector Machines (SVM)

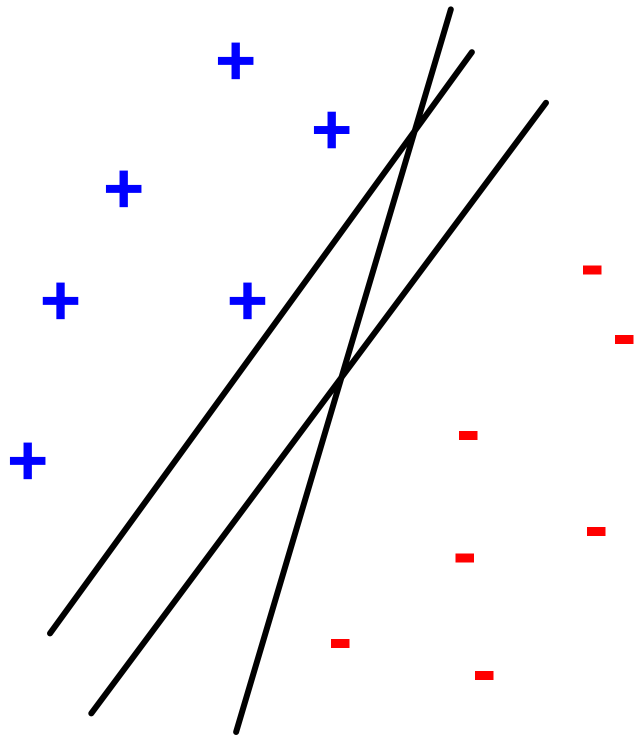
Minor Gordon

gordom6@rpi.edu

Rensselaer

Support Vector Machines

- Want to separate "+" from "-" using a line



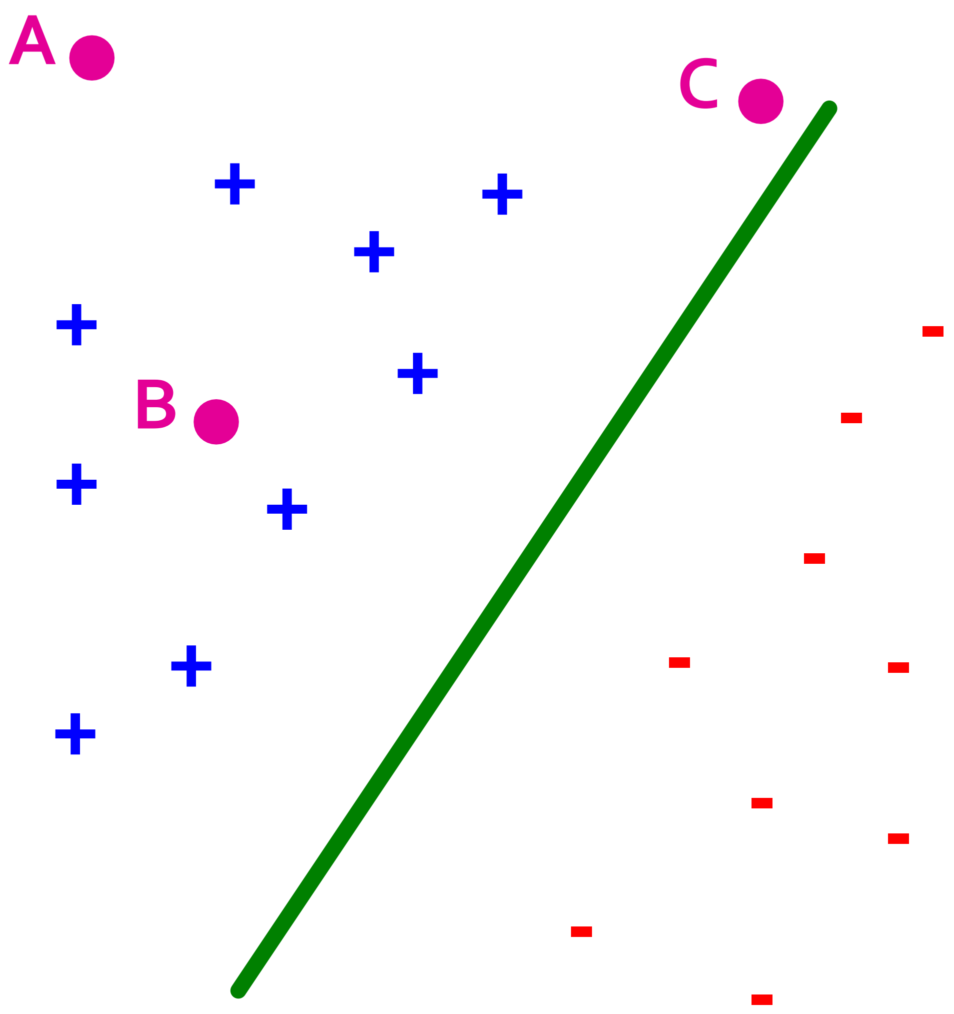
Data:

- **Training examples:**
  - $(x_1, y_1) \ldots (x_n, y_n)$
- **Each example *i*:**
  - $x_i = (x_i^{(1)},\ldots , x_i^{(d)})$
    - $x_i^{(j)}$ is real valued
  - $y_i \in \{ -1, +1 \}$
- **Inner product:**

$$\boldsymbol{w} \cdot \boldsymbol{x} = \sum_{j=1}^{d} w^{(j)} \cdot x^{(j)}$$

Which is best linear separator (defined by *w*)?

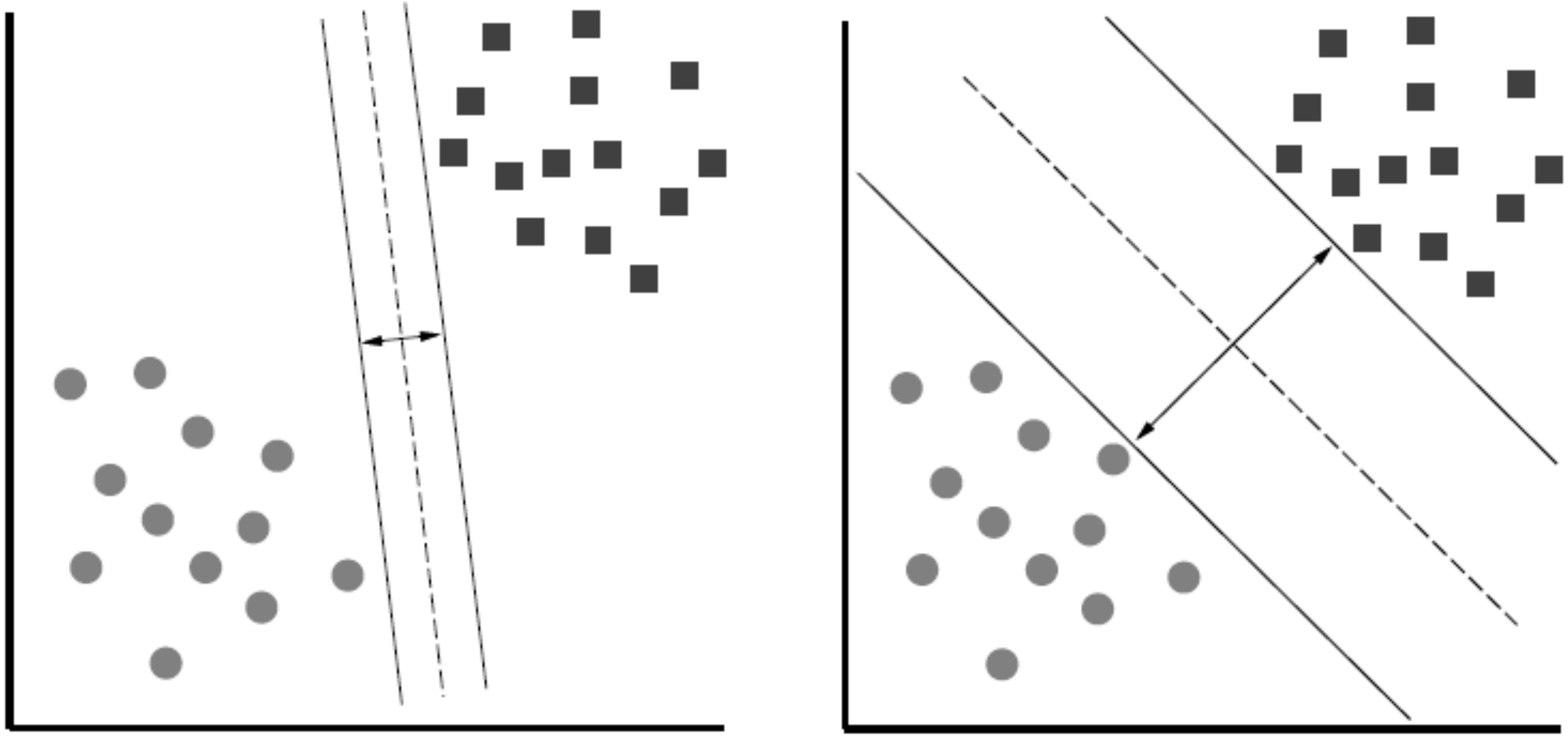
Largest Margin



- **Distance from the separating hyperplane corresponds to the "confidence" of prediction**
- **Example:**
  - We are more sure about the class of **A** and **B** than of **C**

Largest Margin

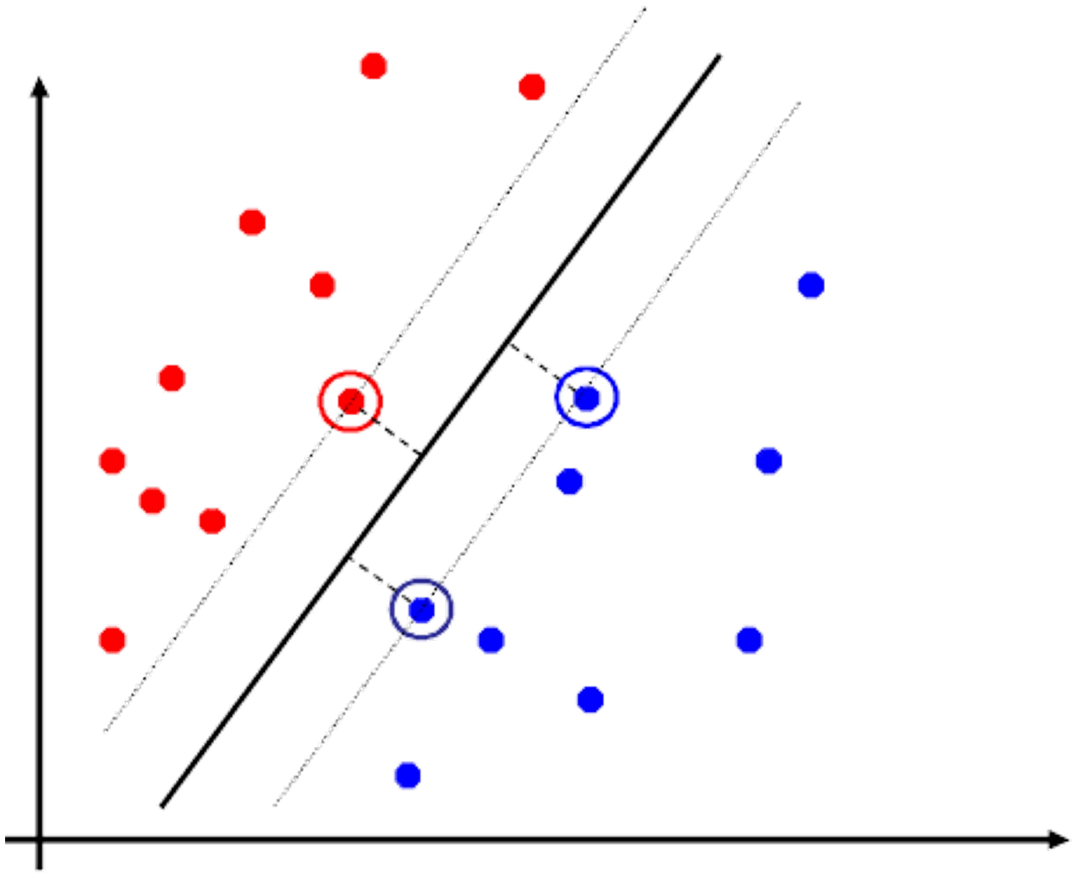
- **Margin γ: Distance of closest example from the decision line/hyperplane**



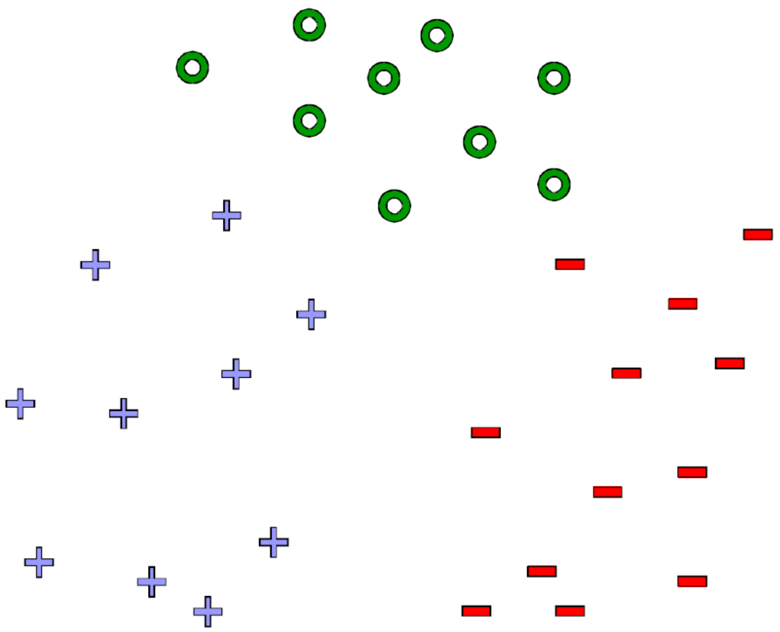
The reason we define margin this way is due to theoretical convenience and existence of generalization error bounds that depend on the value of margin.

Support Vector Machines

- **Separating hyperplane is defined by the support vectors**
  - Points on **+/-** planes from the solution
  - If you knew these points, you could ignore the rest
  - Generally, $d+1$ support vectors (for d dim. data)



What about multiple classes?



- **Idea 1:**
  One against all
  Learn 3 classifiers
  - **+ vs. {o, -}**
  - **- vs. {o, +}**
  - **o vs. {+, -}**

  Obtain:

  $\mathbf{w_+ \, b_+, \;\; w_- \, b_-, \;\; w_o \, b_o}$
- **How to classify?**
- Return class *c*
  $\arg\max_c \; w_c x + b_c$

Learn 1 classifier: Multiclass SVM

- **Idea 2: Learn 3 sets of weights simultaneously!**
  - For each class ***c*** estimate $\boldsymbol{w_c}$, $\boldsymbol{b_c}$
  - **Want the correct class to have highest margin:**

$$\mathbf{w_{y_i}}\,\mathbf{x_i} + \mathbf{b_{y_i}} \geq 1 + \mathbf{w_c}\,\mathbf{x_i} + \mathbf{b_c} \quad \forall \mathbf{c} \neq \mathbf{y_i}\ , \forall \mathbf{i}$$