Non linearly separable data?

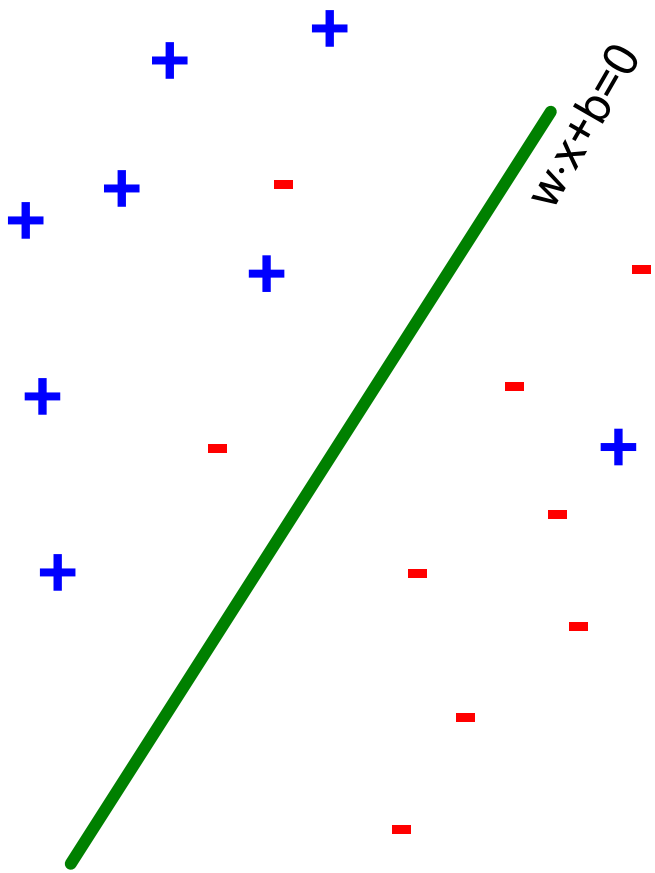
- **If data is not separable introduce penalty:**

$$\min_w \frac{1}{2}\|w\|^2 + C \ (\#\text{number of mistakes})$$

$$s.t. \forall i, y_i(w \cdot x_i + b) \geq 1$$

  - Minimize $\|w\|^2$ plus the number of training mistakes
  - Set ***C*** using cross validation

- **How to penalize mistakes?**
  - **All mistakes are not equally bad!**



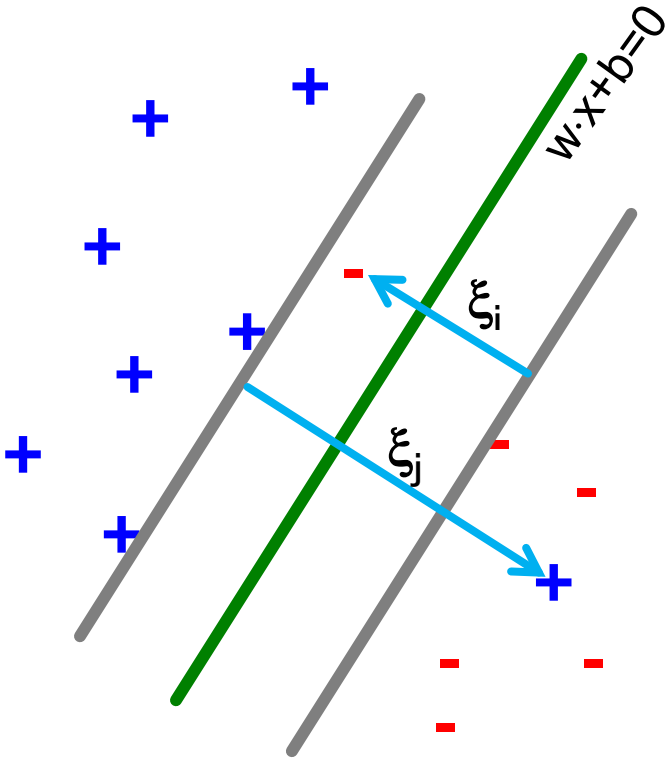
Support Vector Machines

- **Introduce slack variables ξ_i**

$$\min_{w,b,\xi_i \geq 0} \ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i$$

$$s.t. \forall i, y_i(w \cdot x_i + b) \geq 1 - \xi_i$$

- If point $\boldsymbol{x_i}$ is on the wrong side of the margin then get penalty ξ_i



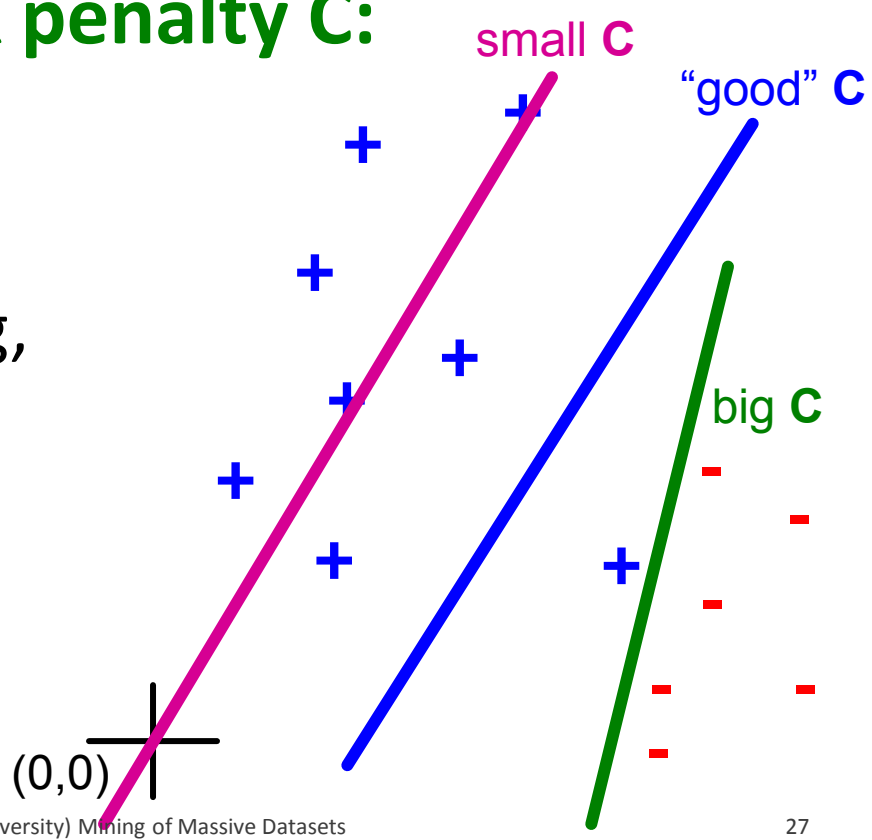
For each datapoint:
If margin ≥ 1, don't care
If margin < 1, pay linear penalty

Slack Penalty C

$$\min_w \tfrac{1}{2}\|w\|^2 + \mathrm{C}\ \ (\#\text{number of mistakes})$$

$$s.t.\ \forall i, y_i(w \cdot x_i + b) \geq 1$$

- **What is the role of slack penalty C:**
  - **C=∞:** Only want to **w, b** that separate the data
  - **C=0:** Can set ξ_i to anything, then **w=0** (basically ignores the data)

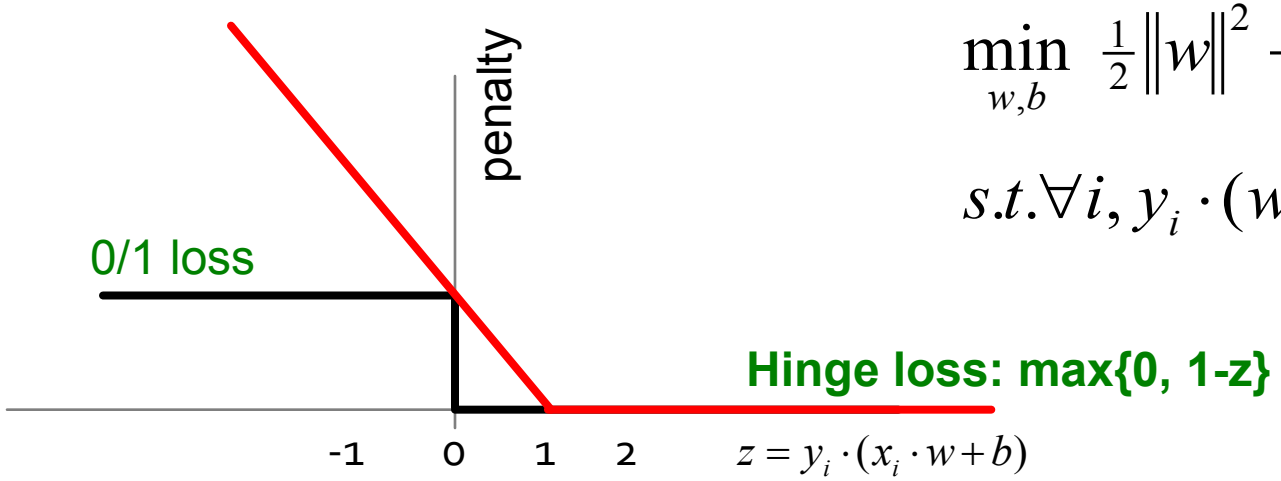


Support Vector Machines

- **SVM in the "natural" form**

$$\arg\min_{w,b} \; \tfrac{1}{2} \underbrace{w \cdot w}_{\text{Margin}} + \underbrace{C}_{\text{Regularization parameter}} \sum_{i=1}^{n} \underbrace{\max\{0, 1 - y_i(w \cdot x_i + b)\}}_{\text{Empirical loss L (how well we fit training data)}}$$

- **SVM uses "Hinge Loss":**



$$\min_{w,b} \; \tfrac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i$$

$$s.t. \forall i, \; y_i \cdot (w \cdot x_i + b) \geq 1 - \xi_i$$