Applied Machine Learning

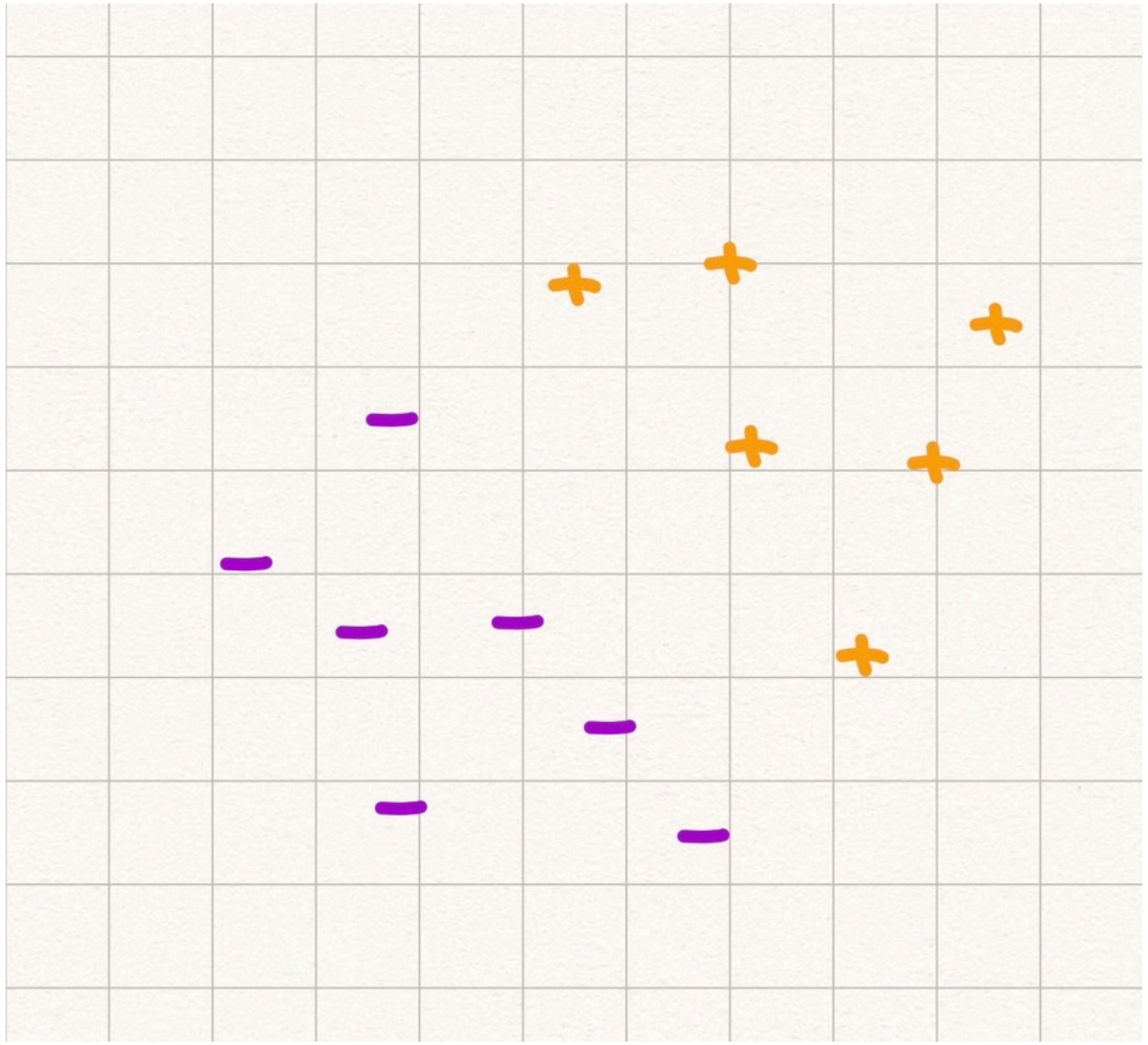
Classification - Support Vector Machines

Support Vector Machine

- Overview
- Training Error Cost - Hinge Loss
- Regularization Constant

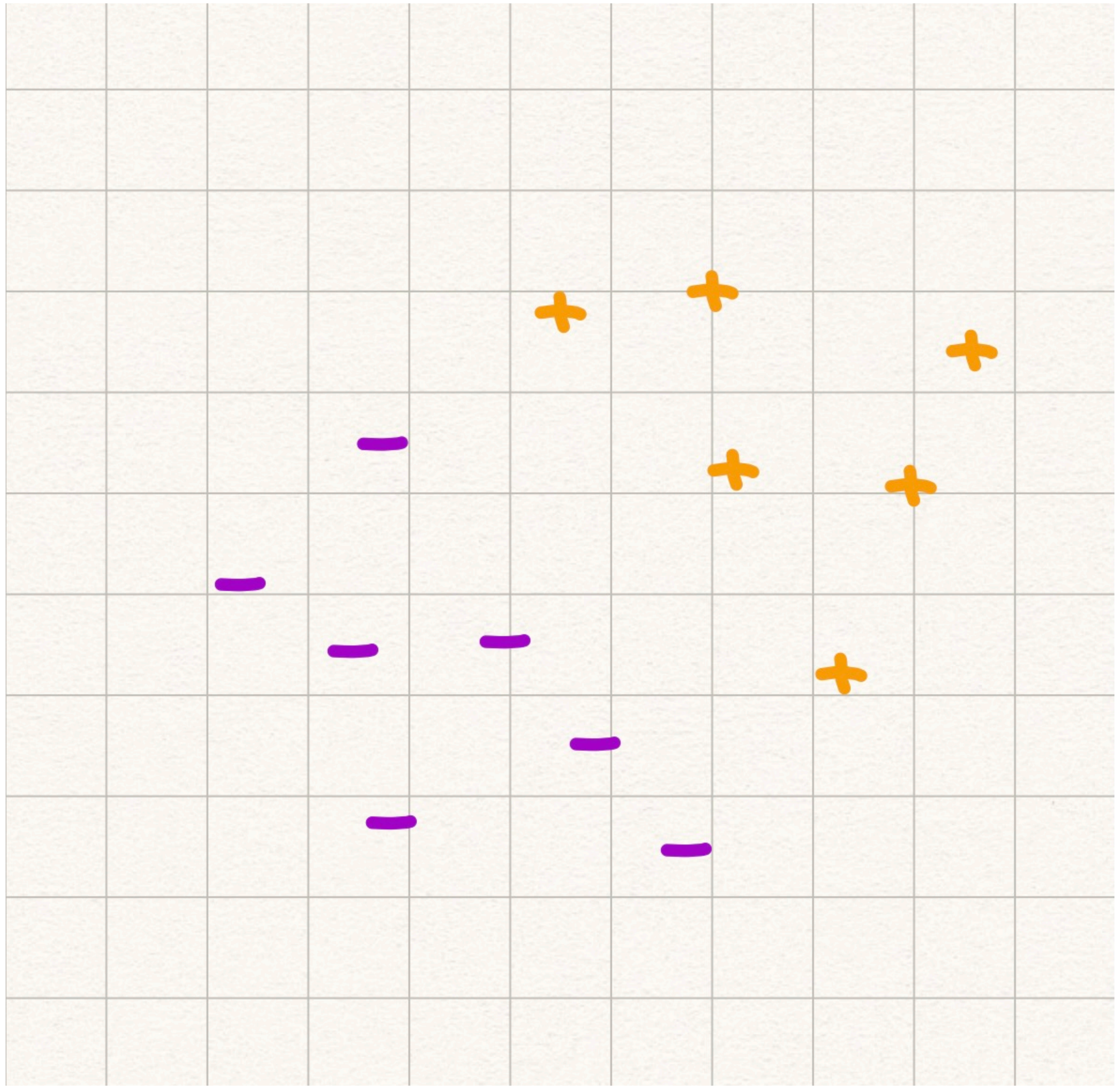
SVM - Overview

- Binary linear classifier
  - $Class \in \{+1, -1\}$
- Easy to train
- Fast classification
- Hinge Loss
  - $Cost$(prediction, true label)



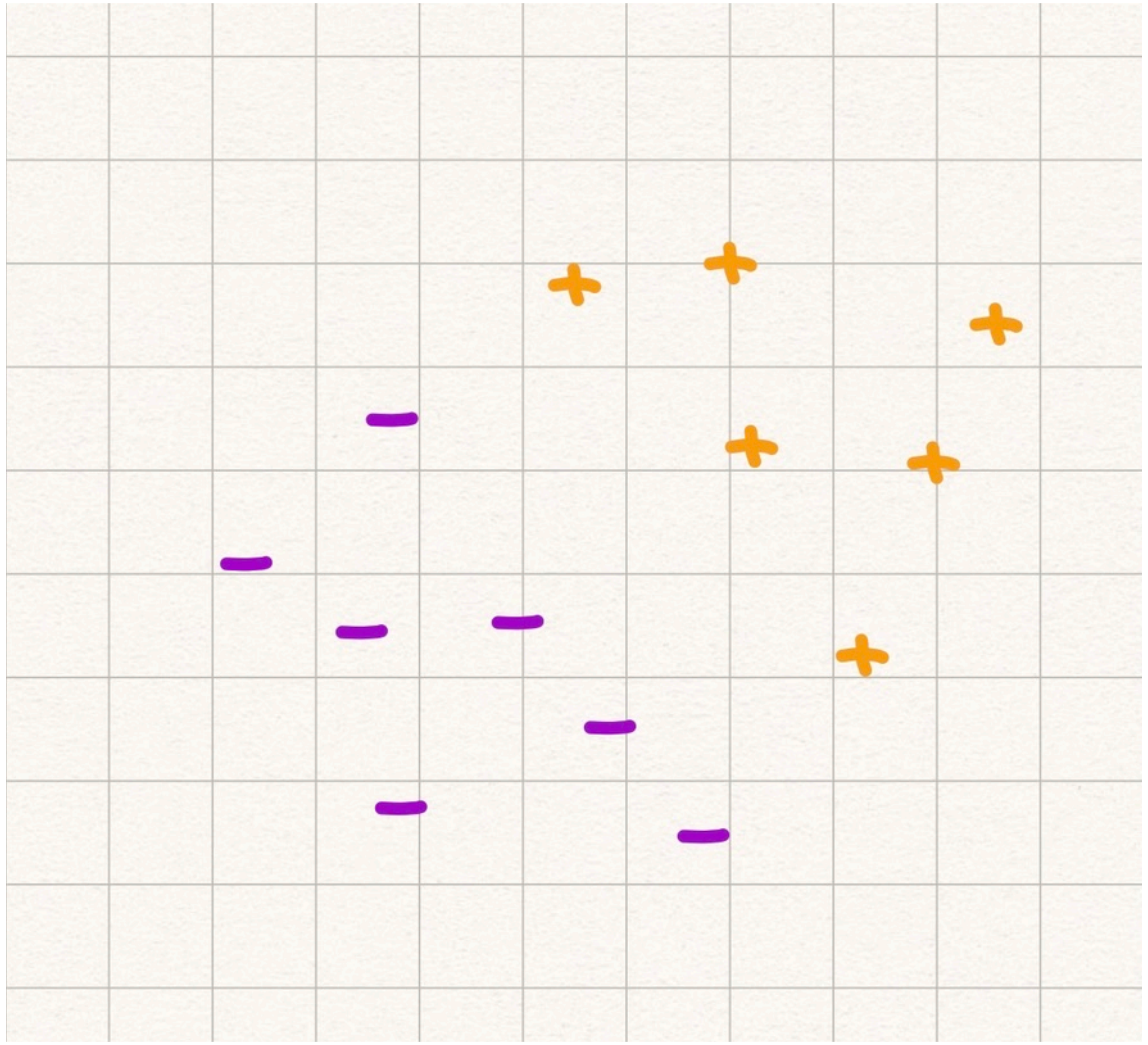
SVM - Overview

- Real-valued features
- No missing feature values
- Classes separable through a linear function
- Need enough features



SVM - Overview

- $Class \in \{+1, -1\}$
- Decision Boundary:
    - $\mathbf{a}^\top \mathbf{x} + b$
- Classification:
    - $f(\mathbf{x}) = sign(\mathbf{a}^\top \mathbf{x} + b)$
- Cost function
    - Training error cost + λpenalty



$f(\mathbf{x})$: 1 feature

$$
\begin{aligned}
f(\mathbf{x}) &= sign(\mathbf{a}^\top \mathbf{x} + b) \\
&= sign(ax + b)
\end{aligned}
$$

Boundary

$$
\begin{aligned}
ax + b &= 0 \\
x &= -\frac{b}{a}
\end{aligned}
$$

Classification

$$
y = \begin{cases} 1 & \text{if } x \geq -\frac{b}{a} \\ -1 & \text{if } x < -\frac{b}{a} \end{cases}
$$

$f(\mathbf{x})$: 2 features

$$f(\mathbf{x}) = sign(\mathbf{a}^\top \mathbf{x} + b)$$

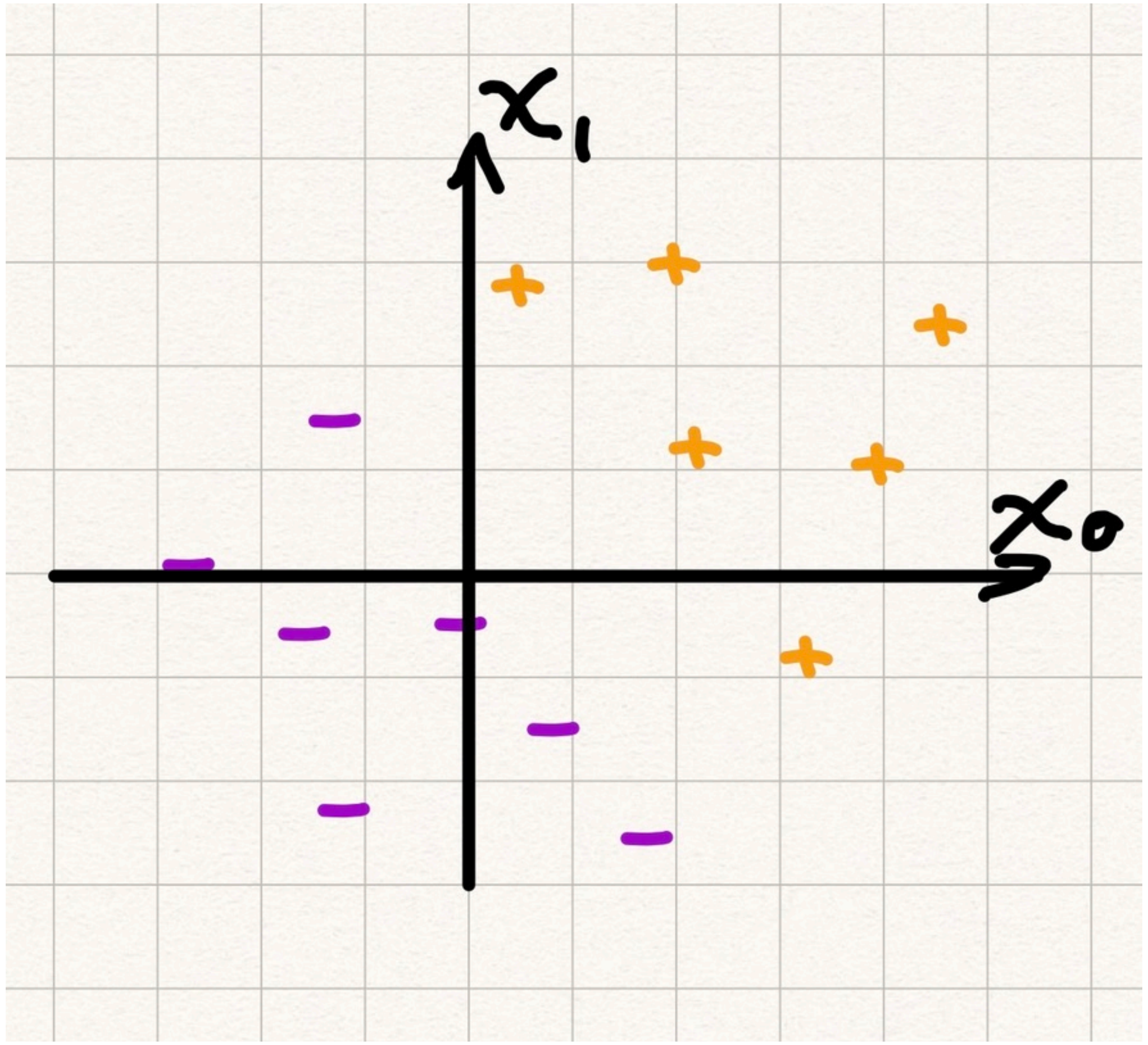
$$\mathbf{x} = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} \qquad \mathbf{a} = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} \qquad b$$

Boundary

$$\begin{aligned} [a_0 \quad a_1] \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} + b &= 0 \\ a_0 x_0 + a_1 x_1 + b &= 0 \\ x_1 &= -\frac{a_0}{a_1} x_0 - \frac{b}{a_1} \end{aligned}$$

Classification

$$y = \begin{cases} 1 & \text{if } x_1 \geq -\frac{a_0}{a_1} x_0 - \frac{b}{a_1} \\ -1 & \text{if } x_1 < -\frac{a_0}{a_1} x_0 - \frac{b}{a_1} \end{cases}$$



$f(\mathbf{x})$: k features

Training with N pairs $(\mathbf{x}_i, y_i)$ to find a and b

Classification

$$f(\mathbf{x}) \quad = \quad sign(\mathbf{a}^\top \mathbf{x} + b)$$

$$\mathbf{x} = \begin{bmatrix} x_0 \\ \vdots \\ x_{k-1} \end{bmatrix} \qquad \mathbf{a} = \begin{bmatrix} a_0 \\ \vdots \\ a_{k-1} \end{bmatrix} \qquad b$$

Cost function

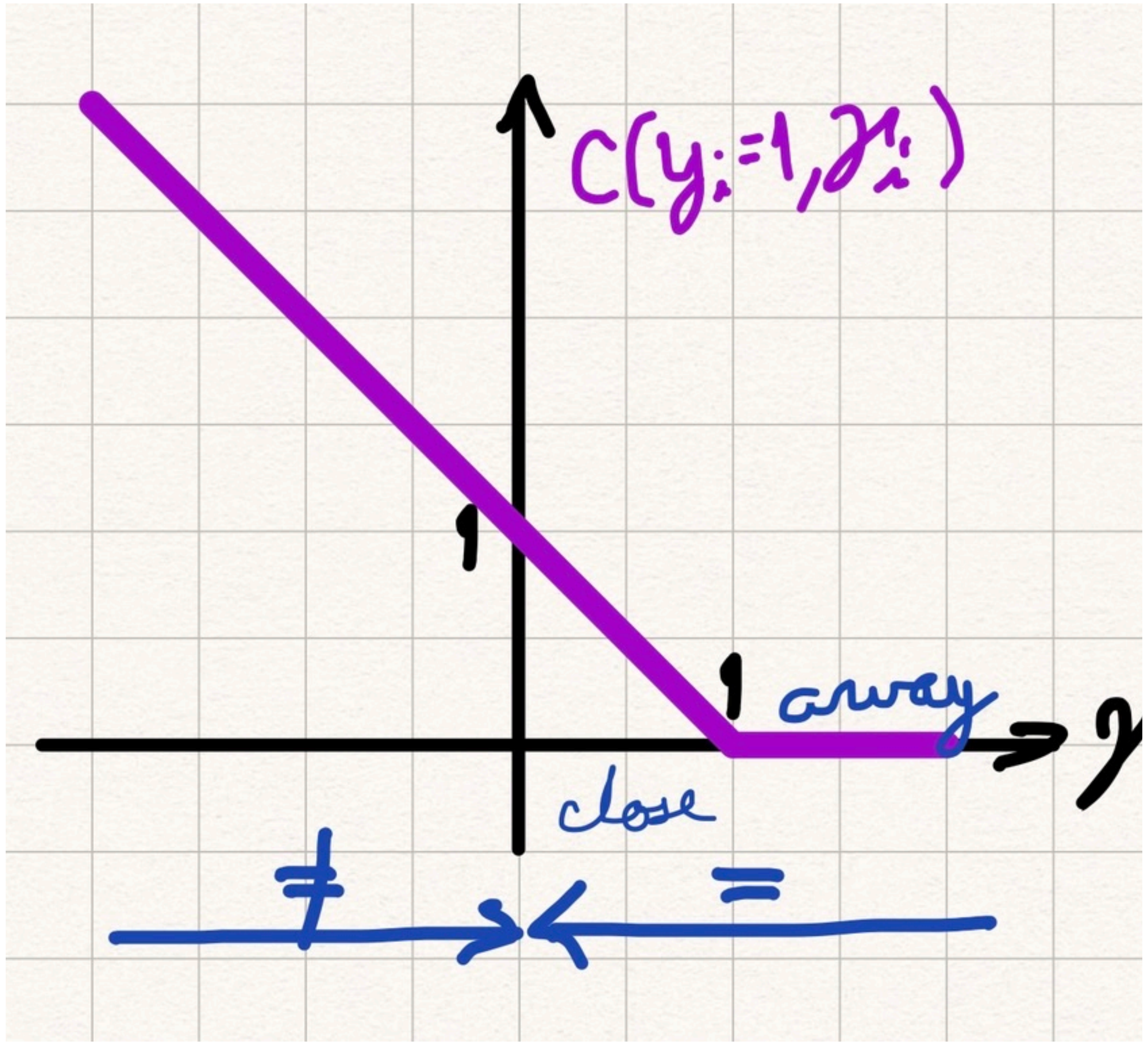
Training error cost + λpenalty

Training cost

- Classification of example i
  - $\gamma_i = \mathbf{a}^\top \mathbf{x}_i + b$
  - $sign(\gamma_i)$
- Cost of classification: $C(y_i, \gamma_i)$
  - $y_i \neq sign(\gamma_i)$: incorrect: $C(y_i, \gamma_i)$ large
  - $y_i = sign(\gamma_i)$: correct:
    - γ_i close: $C(y_i, \gamma_i)$ medium
    - γ_i away: $C(y_i, \gamma_i)$ no cost

Training cost: Hinge Loss

- Classification of example i
  - $\gamma_i = \mathbf{a}^\top \mathbf{x}_i + b$
  - $sign(\gamma_i)$
- Cost of classification: $C(y_i, \gamma_i) = \max(0, 1 - y_i * \gamma_i)$
  - $y_i \neq sign(\gamma_i)$: incorrect: $C(y_i, \gamma_i) = |\gamma_i| + 1$: large
  - $y_i = sign(\gamma_i)$: correct:
    - $\gamma_i < 1$ close: $C(y_i, \gamma_i) = |\gamma_i| + 1$: medium
    - $\gamma_i \geq 1$ away: $C(y_i, \gamma_i) = 0$: no cost



Finding $\mathbf{a}$ and b

- find $\mathbf{a}$ and b that minimize Hinge Loss cost
  - all training examples at the right side of boundary
    - Training error:

$$\frac{1}{N}\sum_{i=1}^{N} C(y_i, \gamma_i) \quad = \quad \frac{1}{N}\sum_{i=1}^{N} \max(0, 1 - y_i * \gamma_i)$$
$$= \quad \frac{1}{N}\sum_{i=1}^{N} \max(0, 1 - y_i * (\mathrm{a}^\top \mathbf{x} + b))$$

  - margin around boundary to reduce errors on new examples

Finding $\mathbf{a}$, b, and λ

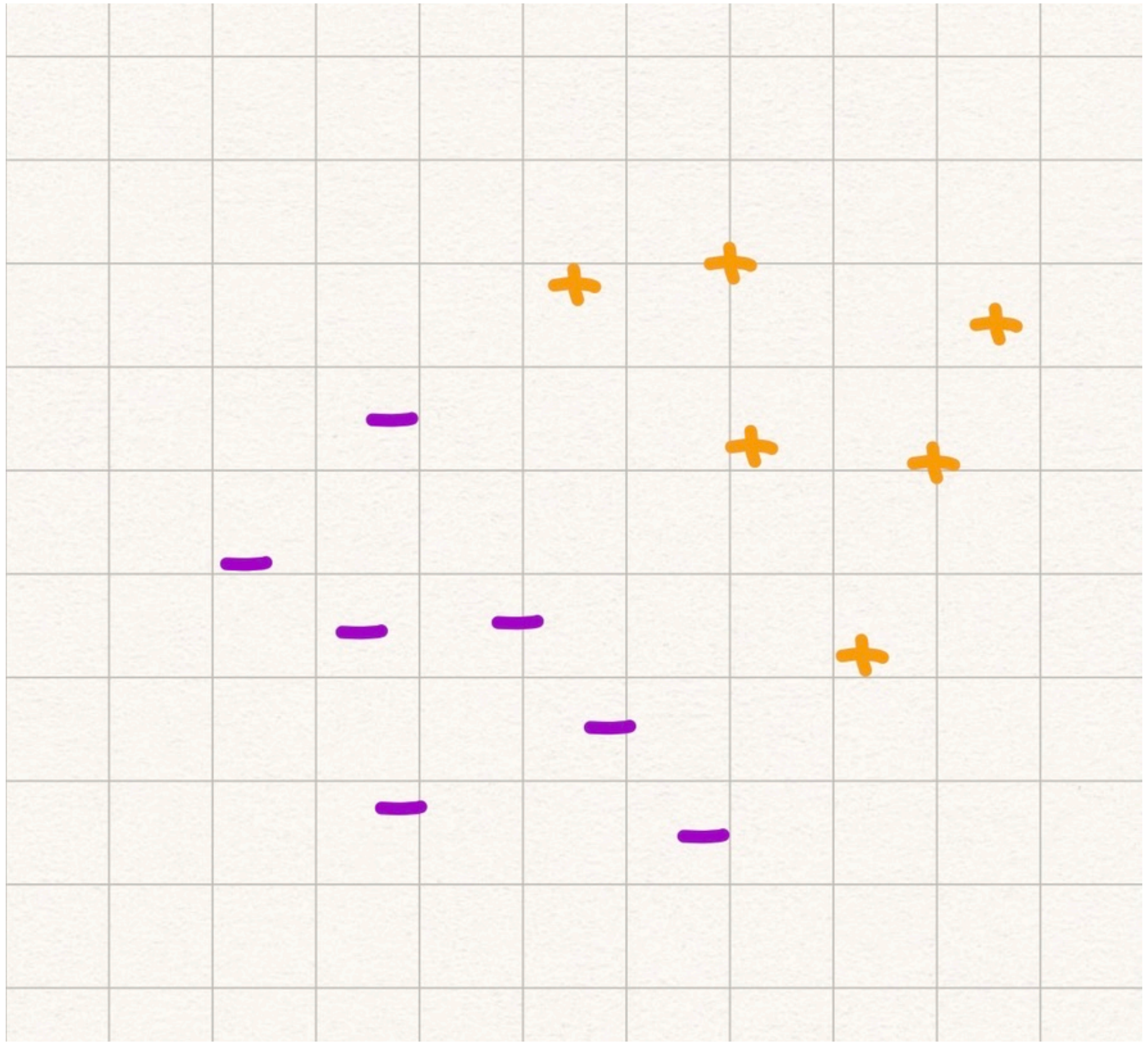
- find $\mathbf{a}$ and b that minimize Hinge Loss cost
    - all training examples at the right side of boundary
    - margin around boundary to reduce errors on new examples
        - Correct classification:
            - hinge loss: 0 away from the boundary, [0-1], close to the boundary
        - Incorrect classification
            - hinge loss: > 1, far from the boundary
            - keep $||\mathbf{a}||$ low: penalize large values of $\frac{1}{2}\mathbf{a}^\top\mathbf{a}$
        - Regularization term:
            - $\lambda\frac{1}{2}\mathbf{a}^\top\mathbf{a}$

Finding $\mathbf{a}$ and b

- find $\mathbf{a}$ and b that minimize Hinge Loss cost
    - all training examples at the right side of boundary
    - margin around boundary to reduce errors on new examples
- Cost Function
    - $$S(\mathbf{a}, b; \lambda) = \frac{1}{N} \sum_{i=1}^{N} [\max(0, 1 - y_i * (\mathbf{a}^\top \mathbf{x} + b))] + \lambda \frac{1}{2} \mathbf{a}^\top \mathbf{a}$$

SVM - Overview

- Easy to train
- Fast classification
- adding features may allow to apply it



Support Vector Machines

- Overview
- Training Error Cost - Hinge Loss
- Regularization Constant

Applied Machine Learning

Classification - Support Vector Machines