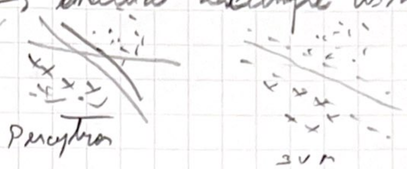
- Uses in cases with multiple classes

Prob.
- Prediction
  - → SVM doesn't have naturally
  - → Perceptron → by adding a activation function

- Main goal of SVM is to separate data as best as possible
  - ↳ include example with draw



- Kernel trick → SVM
  - → more efficient since we don't need to ~~use~~ transform samples of data into a higher dimension

7.

$$g(\alpha) = \tfrac{1}{2}\,\alpha^T Q \alpha + \alpha^T \mathbb{1}_N$$

$$= \underbrace{-\tfrac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j x_i^T x_j}_{1} + \underbrace{\sum_{i=1}^{N} \alpha_i}_{2}$$

1 $$-\tfrac{1}{2} \underbrace{\sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j}\cdot (x_i^T x_j)$$

$$= -\tfrac{1}{2} \cdot \sum_{i=1}^{N} y_i \alpha_i x_i \left(y_1 \alpha_1 x_1 + y_2 \alpha_2 x_2 + \dots + y_N \alpha_N x_N \right)$$

$$= -\tfrac{1}{2} \left(y_1 \alpha_1 x_1 + y_2 \alpha_2 x_2 + \dots + y_N \alpha_N x_N \right)^2 \in \mathbb{R}^{1\times 1}$$

Therefore, given ~~that~~ matrix product is commutative and the dual function SVM format:

$$Q = -y x^T x y^T, \text{ with } x, y \in \mathbb{R}^{N\times 1}$$

b) As seen in the previous exercise (a)), $\alpha^T Q \alpha$ results in a quadratic function (dependent on α). Given the concavity of such function, the local maximizer/minimizer is always a global maximum / minimum.

In order to determine the correct concavity (which must be negative) we need to determine the correct signal of Q, since $\alpha \geq 0$, for $i=1,\dots,N$

Given $Q = \underbrace{(-1)}_{\leq 0} \cdot \underbrace{\sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j x_i x_j}_{\geq 0}$, so $Q \leq 0$

↓

negative semidefinite matrix