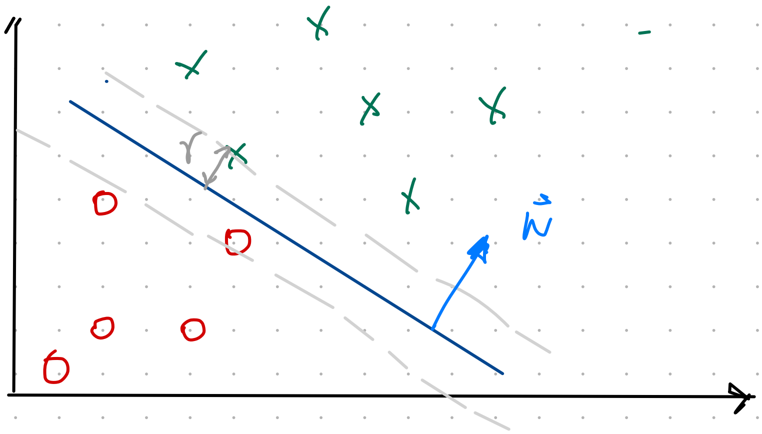
Lecture 21

Support Vector Machine (SVM)

- Idea: Find the maximum margin hyperplane



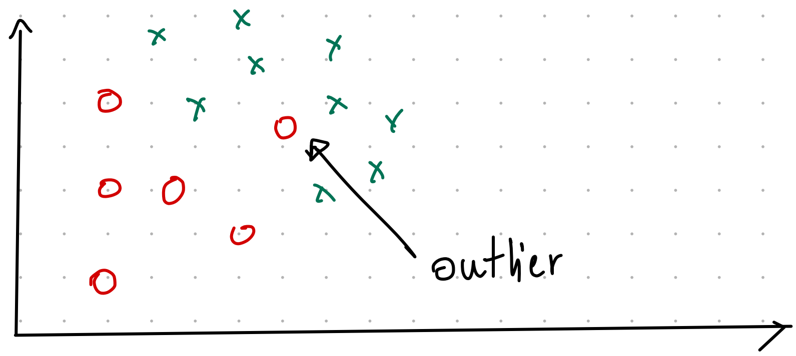
- The SVM's optimization problem: the objective is

$$\underbrace{\min_{\vec{w}, b} \; w^T w}_{\text{quadratic objective function}} \quad \text{s.t.} \quad \forall i \; \underbrace{y_i(\vec{w}^T \vec{x}_i + b) \geq 1}_{\text{linear constraints}}$$

To find $\vec{w}, b$ that minimize the objective, we can use quadratic programming solver (QCQP).

Nice interpretation: Find us the simplest solution such that all points lie at least 1 unit away from the hyperplane on the correct side. The vector $\vec{w}$ (and b) supports the closest points to the hyperplane where $y_i(\vec{w}^T \vec{x}_i + b) = 1$.

- Let's consider the case where data is not linearly separable.



- In this case, no separating hyperplane exists, and that's bad.

- What we would do instead is that we can sacrifice some outlier(s) in order to place the hyperplane.

SVM with Soft Constraints:

- Fix: We allow the constraints to be soften slightly with the introduction of slack variable $\xi_i \geq 0, \forall i$:

(psi) (wolb)

$$\underbrace{\min_{\vec{w}, b, \xi} w^T w + C \sum_{i=1}^{n} \xi_i}_{\text{objective function}} \quad s.t. \quad \underbrace{\forall i \; y_i(\vec{w}^T \vec{x}_i + b) \geq 1 - \xi_i \quad \forall i \; \xi_i \geq 0}_{\text{linear constraints}}$$

- The slack variable ξ_i allow $\vec{x}_i$ to be closer to the hyperplane (or even be on the wrong side) but there in a penalty in the objective function for such slack.

Penalty: $C \to +\infty$, SVM will try to make all the points to be on the right side

$C \to 0$, SVM may sacrifice some points to obtiaed a simpler hyperplane.

- Unconstrainted Formulation: We set ξ_i as followed

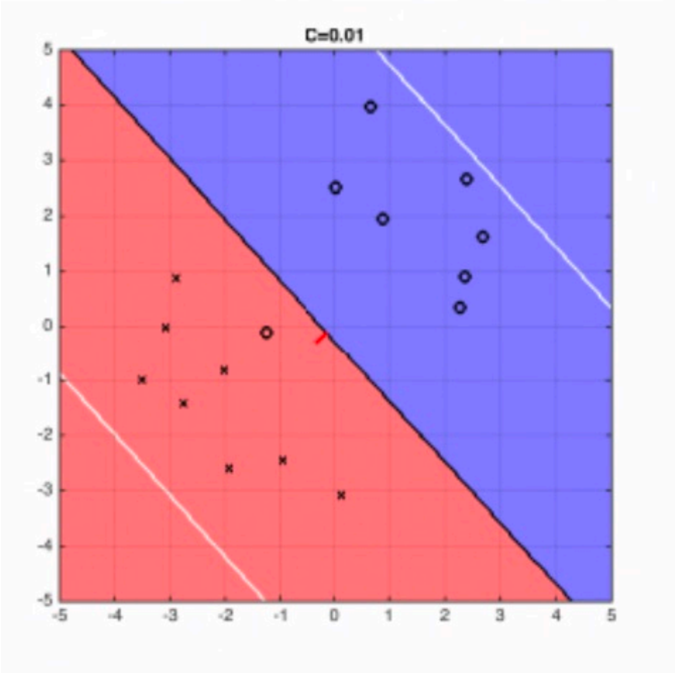
$$\xi_i = \begin{cases} 1 - y_i(\vec{w}^T \vec{x}_i + b) & \text{if } y_i(\vec{w}^T \vec{x}_i + b) < 1 \\ 0 & \text{if } y_i(\vec{w}^T \vec{x}_i + b) \geq 1 \end{cases}$$

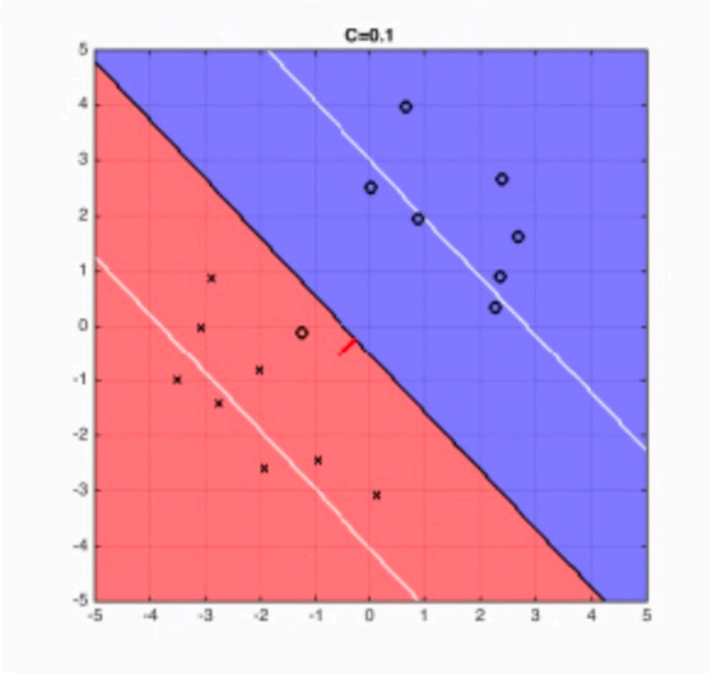
$$\Longleftrightarrow$$

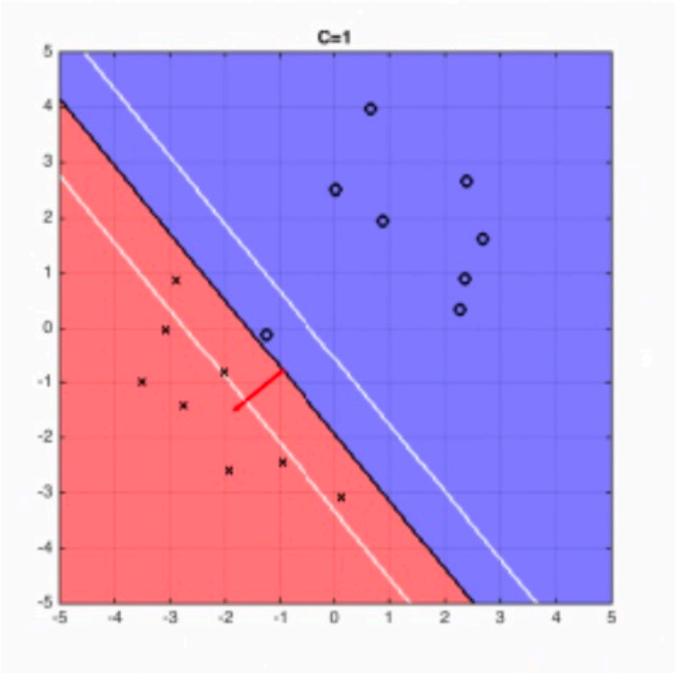
$$\xi_i = \max(1 - y_i(\vec{w}^T \vec{x}_i + b), 0)$$

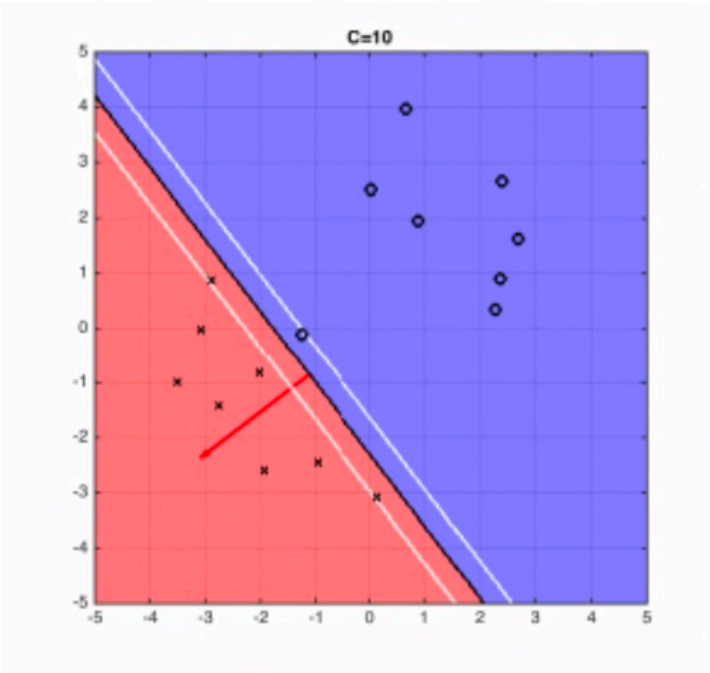
- The SVM with soft constraints optimization problem:

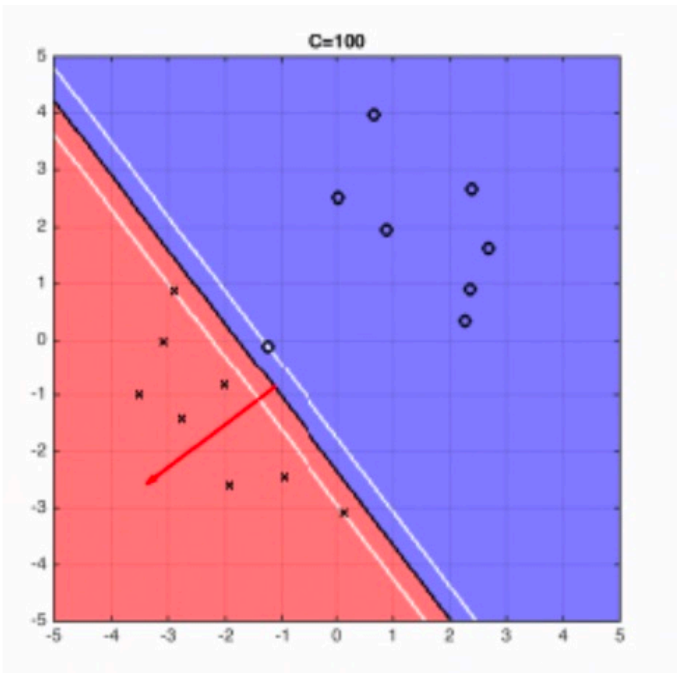
$$\min_{\vec{w}, b} w^T w + C \sum_{i=1}^{n} \max(1 - y_i(\vec{w}^T x + b, 0)$$


C=0.01
5 4 3 2 1 0 -1 -2 -3 -4 -5
-5 -4 -3 -2 -1 0 1 2 3 4 5


C=0.1
5 4 3 2 1 0 -1 -2 -3 -4 -5
-5 -4 -3 -2 -1 0 1 2 3 4 5


C=1
5 4 3 2 1 0 -1 -2 -3 -4 -5
-5 -4 -3 -2 -1 0 1 2 3 4 5


C=10
5 4 3 2 1 0 -1 -2 -3 -4 -5
-5 -4 -3 -2 -1 0 1 2 3 4 5


C=100
5 4 3 2 1 0 -1 -2 -3 -4 -5
-5 -4 -3 -2 -1 0 1 2 3 4 5