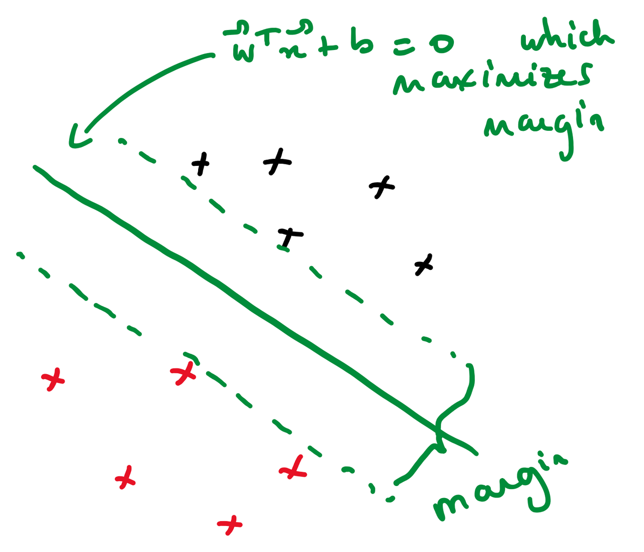
Video 13

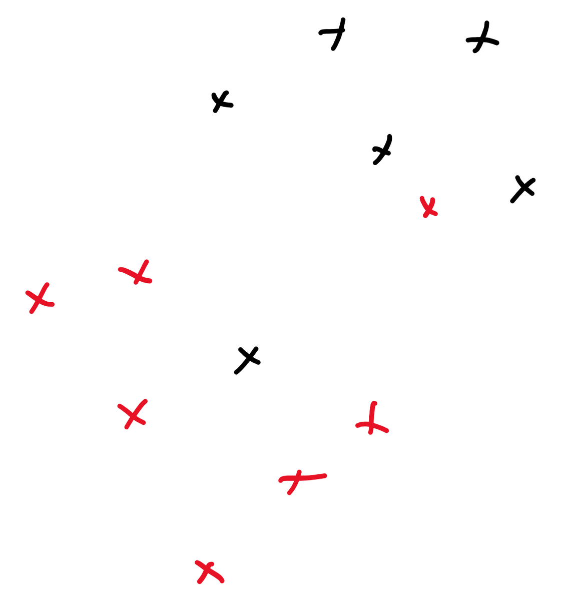
Saturday, October 10, 2020 8:13 PM

SVM

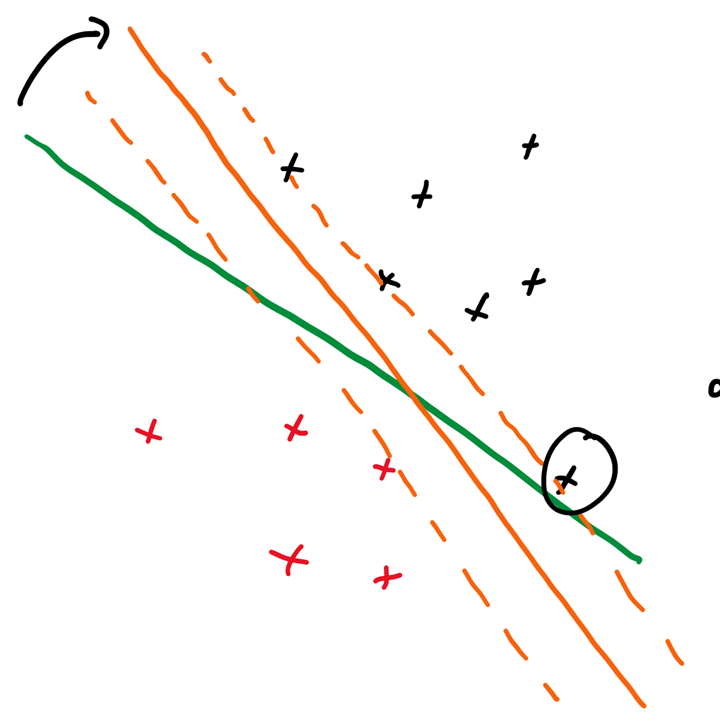


$$\min_{\vec{w},b} \|\vec{w}\|^2 \quad \text{s.t. } t^{(i)}(\vec{w}^T\vec{x}^{(i)}+b) \geq 1$$
$$\text{for all } i \in 1..N$$

How do we apply an SVM on non linearly-separable data?



Data is not linearly separable

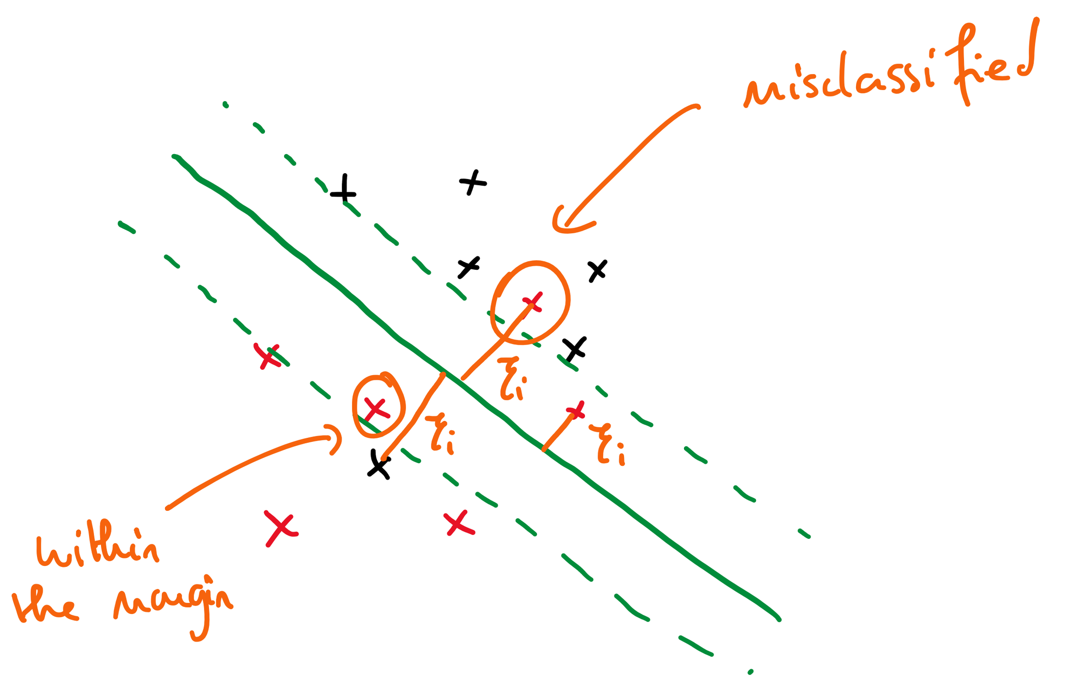


adding this point greatly modifying the max-margin hyperplane.

Data contains outliers

Slack variables: ξ_i

allowing points to be within the margin or even misclassified



Goal: constrain / penalize the total amount of slack.

Soft margin constraint

$$t^{(i)}(\vec{w}^T\vec{x}^{(i)}+b) \geq 1 - \xi_i$$

$$\xi_i \geq 0$$

(as opposed to a hard margin constraint in the first SVM video)

we've added the slack variable to "soften" the margin constraint

Penalize $\sum_{i=1}^{N} \xi_i$

Soft-margin SVM objective: $\min_{\vec{w},b,\xi_i} \frac{1}{2}\|\vec{w}\|^2 + \gamma \sum_{i=1}^{N} \xi_i$ s.t. $t^{(i)}(\vec{w}^T\vec{x}^{(i)}+b) \geq 1-\xi_i$

$\xi_i \geq 0$

(for all $i \in 1..N$)

new hyperparameter which trades off the margin with the amount of slack.

⚠ This a <u>different</u> γ than we used in the first SVM video where γ referred to the distance between a point and the hyperplane.

$\gamma = 0 \longrightarrow$ we will get $\vec{w}=0$ as the solution.

(because we allow ∞ slack, we can misclassify everything, so $\min \frac{1}{2}\|\vec{w}\|^2$ is achieved for $\vec{w}=0$)

$\gamma \longrightarrow \infty \longrightarrow$ we get the hard margin.

Soft margin constraints: $t^{(i)}(\vec{w}^T\vec{x}+b) \geq 1-\xi_i \iff \xi_i \geq 1 - t^{(i)}(\vec{w}^T\vec{x}+b)$

$\xi_i \geq 0$

Case 1: $1 - t^{(i)}(\vec{w}^T\vec{x}+b) \leq 0$

$t^{(i)}(\vec{w}^T\vec{x}+b) \geq 1$

We can set $\xi_i = 0$

Here, we have a correctly classified training example. It's "beyond" the decision boundary & outside the margin

Case 2: $1 - t^{(i)}(\vec{w}^T\vec{x}+b) > 0$

Here, the point is within the margin or misclassified

The smallest ξ_i is $\xi_i = 1 - t^{(i)}(\vec{w}^T\vec{x}+b)$

because $\underbrace{\xi_i \geq 1 - t^{(i)}(\vec{w}^T\vec{x}+b)}_{=} > 0$

If we take both cases, $\boxed{\xi_i = \max\{0, 1 - t^{(i)}(\vec{w}^T\vec{x}+b)\}}$

both constraints on ξ are satisfied.

Soft-margin SVM objective: $\min_{\vec{w},b} \frac{1}{2}\|\vec{w}\|^2 + \gamma \sum_{i=1}^{N} \underbrace{\max\{0, 1 - t^{(i)}(\vec{w}^T\vec{x}+b)\}}_{\xi_i}$