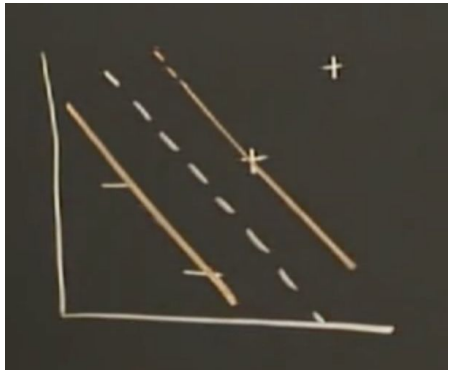
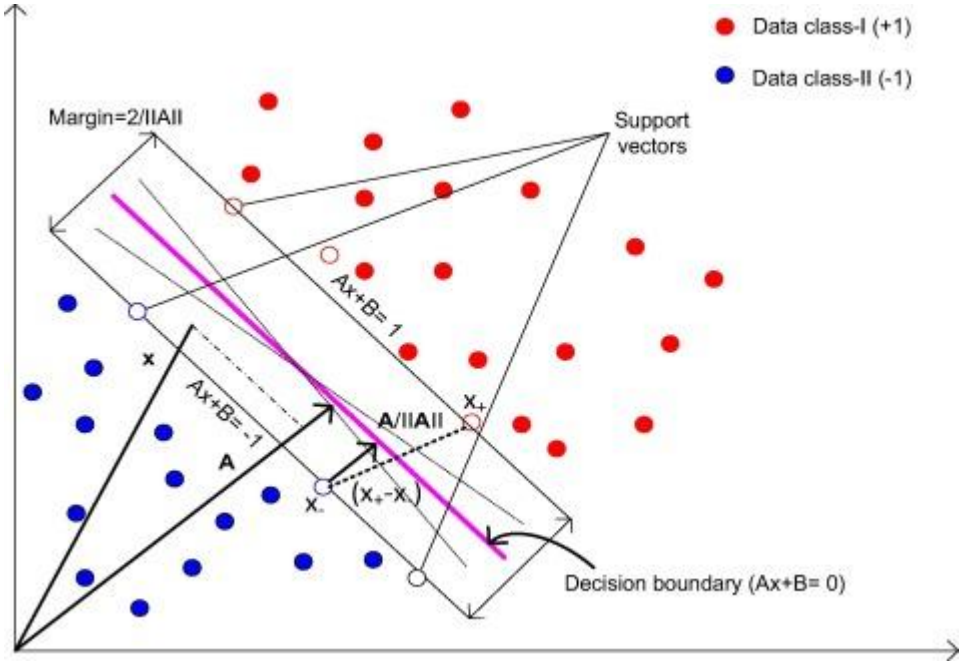
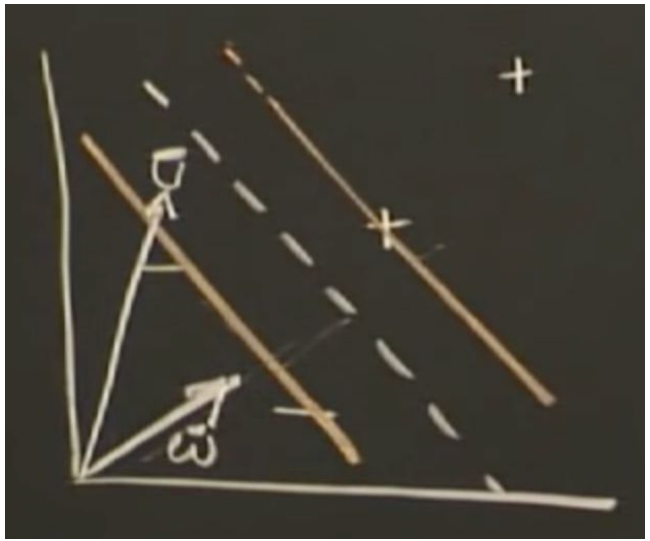
Support Vector Machines

Widest Street Approach





What rule would use this decision boundary?



$$\vec{\omega} \cdot \vec{u} + b \geq 0$$

Vector perpendicular to decision boundary

Position vector of element to classify

Our decision rule!

Minimize

$$\left[\frac{1}{n}\sum_{i=1}^{n}\max\left(0,1-y_i(w\cdot x_i-b)\right)\right]+\lambda\|w\|^2.$$

Linearly Inseparable?

Transform into a higher dimension! Dot products of the vectors still apply

Susceptible to overfitting

https://youtu.be/_PwhiWxHK8o?t=40m43s - failing to linearly separate

https://youtu.be/_PwhiWxHK8o?t=44m51s - works in a higher dimension

In general, it's a separating hyperplane

Good for?

Works only with two categories

Non probabilistic binary linear classifier

Text and hypertext classification

Image classification

Biology - classifying proteins

Stochastic Gradient Descent

Modifies Scaling Gradient Descent

The loss function depends on every data point and you have to calculate on every iteration

Stochastic gradient descent utilizes a terrible estimator in lieu of this, which only sort of works because you make it random