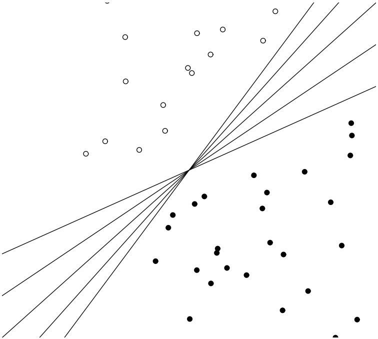
SVM

Data Science: Jordan Boyd-Graber
University of Maryland
SLIDES ADAPTED FROM HINRICH SCHÜTZE

Which hyperplane?



Which hyperplane?

- For linearly separable training sets: there are **infinitely** many separating hyperplanes.
- They all separate the training set perfectly . . .
- . . . but they behave differently on test data.
- Error rates on new data are low for some, high for others.
- How do we find a low-error separator?

Support vector machines

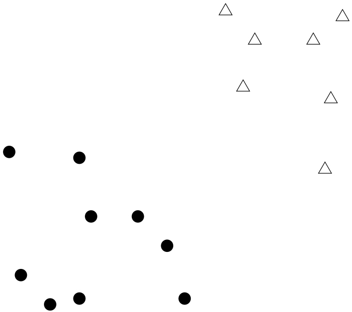
- Machine-learning research in the last two decades has improved classifier effectiveness.
- New generation of state-of-the-art classifiers: support vector machines (SVMs), boosted decision trees, regularized logistic regression, neural networks, and random forests
- Applications to IR problems, particularly text classification

SVMs: A kind of large-margin classifier

Vector space based machine-learning method aiming to find a decision boundary between two classes that is maximally far from any point in the training data (possibly discounting some points as outliers or noise)

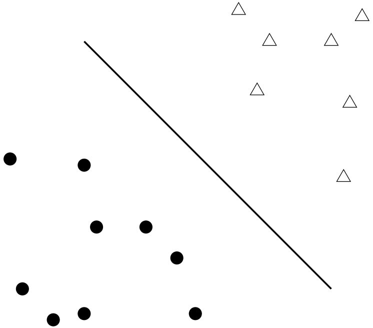
Support Vector Machines

- 2-class training data



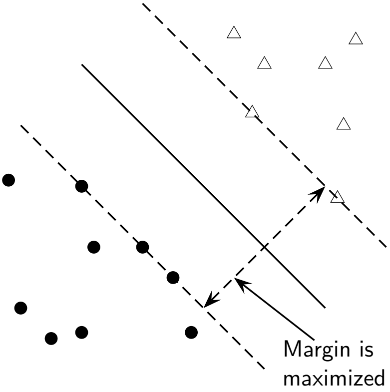
Support Vector Machines

- 2-class training data
- decision boundary → **linear separator**



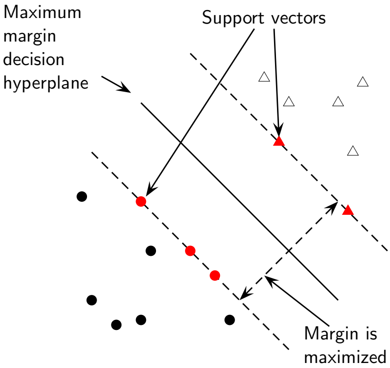
Support Vector Machines

- 2-class training data
- decision boundary ⟶ **linear separator**
- criterion: being maximally far away from any data point ⟶ determines classifier **margin**



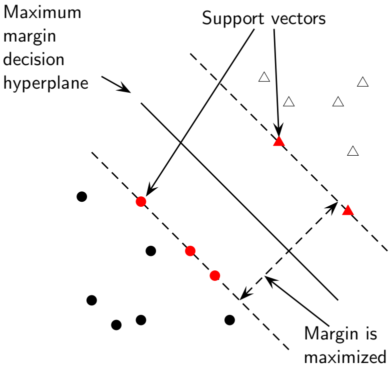
Support Vector Machines

- 2-class training data
- decision boundary ⟶ **linear separator**
- criterion: being maximally far away from any data point ⟶ determines classifier **margin**
- linear separator position defined by **support vectors**



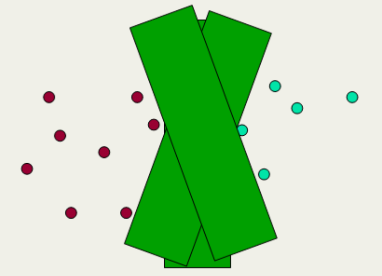
Why maximize the margin?

- Points near decision surface ⟶ uncertain classification decisions
- A classifier with a large margin is always confident
- Gives classification safety margin (measurement or variation)



Why maximize the margin?

- SVM classifier: large margin around decision boundary
- compare to decision hyperplane: place fat separator between classes
  - unique solution
- decreased memory capacity
- increased ability to correctly generalize to test data



Equation

- Equation of a hyperplane

$$\vec{w} \cdot x_i + b = 0 \quad (1)$$

- Distance of a point to hyperplane

$$\frac{|\vec{w} \cdot x_i + b|}{||\vec{w}||} \quad (2)$$

- The margin ρ is given by

$$\rho \equiv \min_{(x,y) \in S} \frac{|\vec{w} \cdot x_i + b|}{||\vec{w}||} = \frac{1}{||\vec{w}||} \quad (3)$$

Equation

- Equation of a hyperplane

$$\vec{w} \cdot x_i + b = 0 \tag{1}$$

- Distance of a point to hyperplane

$$\frac{|\vec{w} \cdot x_i + b|}{||\vec{w}||} \tag{2}$$

- The margin ρ is given by

$$\rho \equiv \min_{(x,y) \in S} \frac{|\vec{w} \cdot x_i + b|}{||\vec{w}||} = \frac{1}{||\vec{w}||} \tag{3}$$

- This is because for any point on the marginal hyperplane, $\vec{w} \cdot x + b = \pm 1$

Optimization Problem

We want to find a weight vector $\vec{w}$ and bias b that optimize

$$\min_{\vec{w},b} \frac{1}{2}||w||^2 \tag{4}$$

subject to $y_i(\vec{w} \cdot x_i + b) \geq 1, \forall i \in [1, m]$.