Department of Computer Science
UNIVERSITY OF COLORADO BOULDER

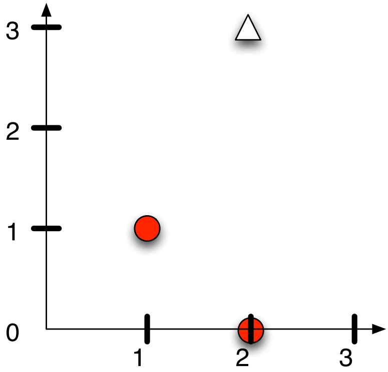
SVM

Introduction to Data Science Algorithms

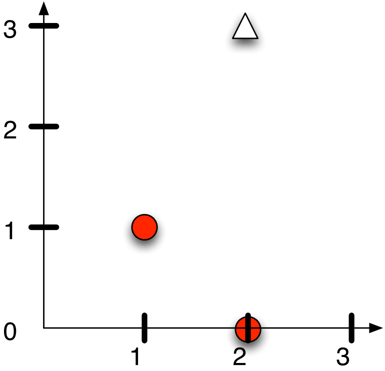
Jordan Boyd-Graber and Michael Paul

SLIDES ADAPTED FROM HINRICH SCHÜTZE

Find the maximum margin hyperplane



Find the maximum margin hyperplane



Which are the support vectors?

Walkthrough example: building an SVM over the data shown

Working geometrically:

Walkthrough example: building an SVM over the data shown

Working geometrically:

- If you got $0 = .5x + y - 2.75$, close!

Walkthrough example: building an SVM over the data shown

Working geometrically:

- If you got $0 = .5x + y - 2.75$, close!
- Set up system of equations

$$w_1 + w_2 + b = -1 \qquad (1)$$

$$\frac{3}{2}w_1 + 2w_2 + b = 0 \qquad (2)$$

$$2w_1 + 3w_2 + b = +1 \qquad (3)$$

Walkthrough example: building an SVM over the data shown

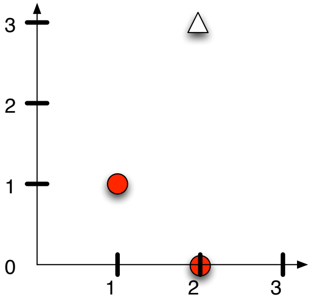
Working geometrically:

- If you got $0 = .5x + y - 2.75$, close!
- Set up system of equations

$$w_1 + w_2 + b = -1 \qquad (1)$$

$$\frac{3}{2}w_1 + 2w_2 + b = 0 \qquad (2)$$

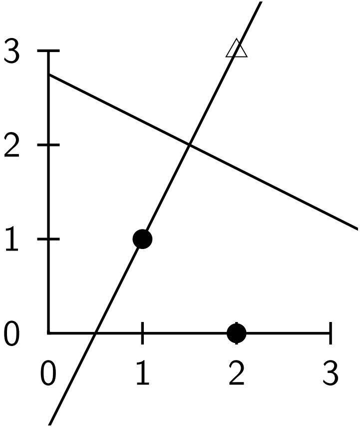
$$2w_1 + 3w_2 + b = +1 \qquad (3)$$



The SVM decision boundary is:

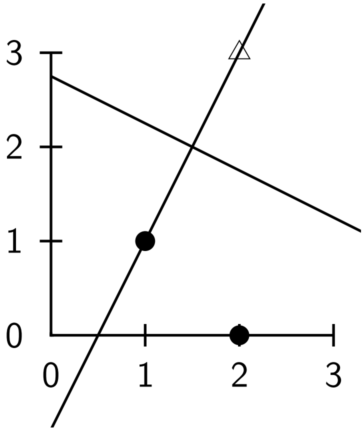
$$0 = \frac{2}{5}x + \frac{4}{5}y - \frac{11}{5}$$

Cannonical Form



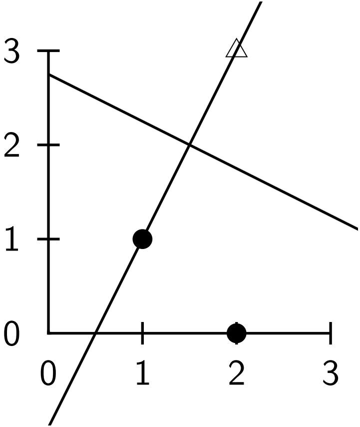
$$w_1 x_1 + w_2 x_2 + b$$

Cannonical Form



$.4x_1 + .8x_2 - 2.2$

Cannonical Form



$.4x_1 + .8x_2 - 2.2$

- $.4 \cdot 1 + .8 \cdot 1 - 2.2 = -1$
- $.4 \cdot \frac{3}{2} + .8 \cdot 2 = 0$
- $.4 \cdot 2 + .8 \cdot 3 - 2.2 = +1$

What's the margin?

- Distance to closest point

What's the margin?

- Distance to closest point

$$\sqrt{\left(\frac{3}{2}-1\right)^2+(2-1)^2}=\frac{\sqrt{5}}{2} \tag{4}$$

What's the margin?

- Distance to closest point

$$\sqrt{\left(\frac{3}{2}-1\right)^2+(2-1)^2}=\frac{\sqrt{5}}{2} \tag{4}$$

- Weight vector

What's the margin?

- Distance to closest point

$$\sqrt{\left(\frac{3}{2}-1\right)^2+(2-1)^2}=\frac{\sqrt{5}}{2} \tag{4}$$

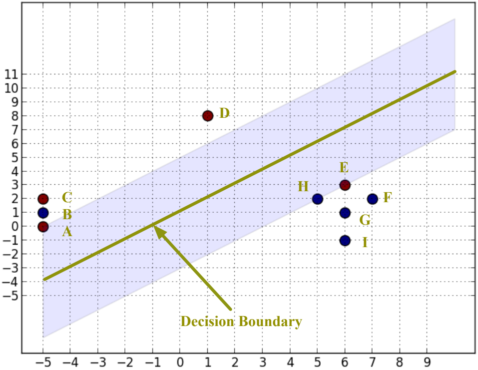
- Weight vector

$$\frac{1}{||w||}=\frac{1}{\sqrt{\left(\frac{2}{5}\right)^2+\left(\frac{4}{5}\right)^2}}=\frac{1}{\sqrt{\frac{20}{25}}}=\frac{5}{\sqrt{5}\sqrt{4}}=\frac{\sqrt{5}}{2} \tag{5}$$

Slack Example

Decision function:

$$w = \begin{bmatrix} -\frac{1}{4} \\ \frac{1}{4} \end{bmatrix}; b = -\frac{1}{4}$$

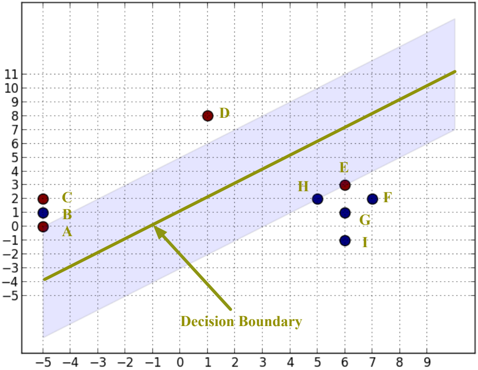


Slack Example

Decision function:

$$w = \begin{bmatrix} -\frac{1}{4} \\ \frac{1}{4} \end{bmatrix}; b = -\frac{1}{4}$$

- What are the support vectors?

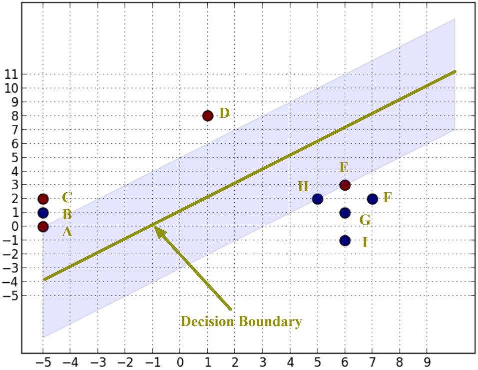


Slack Example

Decision function:

$$w = \begin{bmatrix} -\frac{1}{4} \\ \frac{1}{4} \end{bmatrix}; b = -\frac{1}{4}$$

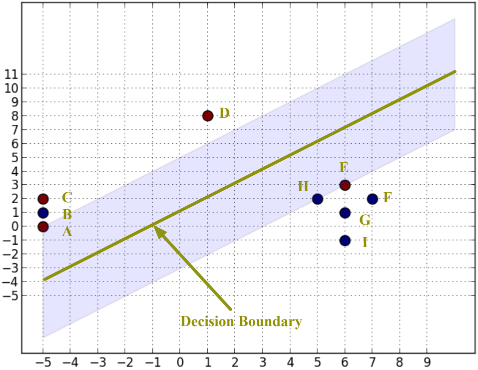
- What are the support vectors?
- Which have non-zero slack?



Slack Example

Decision function:

$$w = \begin{bmatrix} -\frac{1}{4} \\ \frac{1}{4} \end{bmatrix}; b = -\frac{1}{4}$$



- What are the support vectors?
- Which have non-zero slack?
- Compute ξ_B, ξ_E

Computing slack

$$y_i(\vec{w}_i \cdot x_i + b) \geq 1 - \xi_i \tag{6}$$

Computing slack

$$y_i(\vec{w}_i \cdot x_i + b) \geq 1 - \xi_i \tag{6}$$

Point B

$$y_B(\vec{w}_B \cdot x_B + b) = \tag{7}$$

$$-1(-0.25 \cdot -5 + 0.25 \cdot 1 - 0.25) = -1.25 \tag{8}$$

Thus, $\xi_B = 2.25$

Computing slack

$$y_i(\vec{w}_i \cdot x_i + b) \geq 1 - \xi_i \tag{6}$$

Point B

$$y_B(\vec{w}_B \cdot x_B + b) = \tag{7}$$

$$-1(-0.25 \cdot -5 + 0.25 \cdot 1 - 0.25) = -1.25 \tag{8}$$

Thus, $\xi_B = 2.25$

Point E

$$y_E(\vec{w}_E \cdot x_E + b) = \tag{9}$$

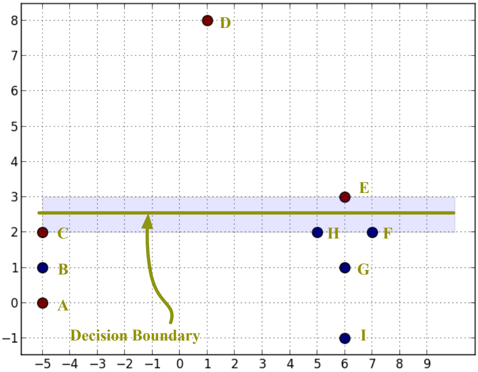
$$1(-0.25 \cdot 6 + 0.25 \cdot 3 + -0.25) = -1 \tag{10}$$

Thus, $\xi_E = 2$

Slack Example

Decision function:

$$w = \begin{bmatrix} 0 \\ 2 \end{bmatrix}; b = -5$$

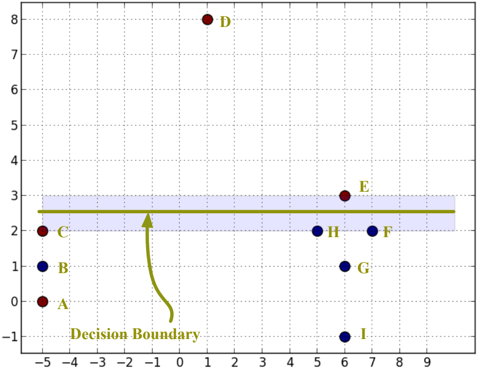


Slack Example

Decision function:

$$w = \begin{bmatrix} 0 \\ 2 \end{bmatrix}; b = -5$$

- What are the support vectors?

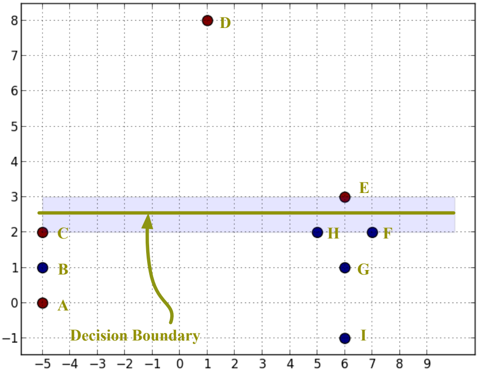


Slack Example

Decision function:

$$w = \begin{bmatrix} 0 \\ 2 \end{bmatrix}; b = -5$$

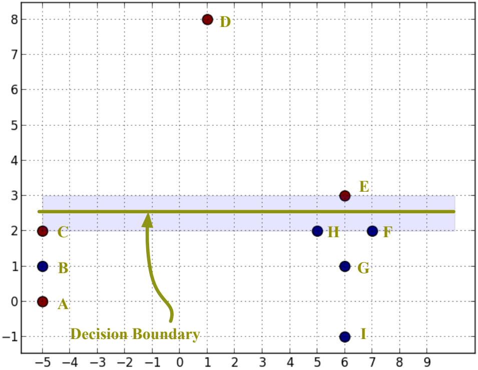
- What are the support vectors?
- Which have non-zero slack?



Slack Example

Decision function:

$$w = \begin{bmatrix} 0 \\ 2 \end{bmatrix}; b = -5$$



- What are the support vectors?
- Which have non-zero slack?
- Compute ξ_A, ξ_C

Computing slack

$$y_i(\vec{w}_i \cdot x_i + b) \geq 1 - \xi_i \quad (11)$$

Computing slack

$$y_i(\vec{w}_i \cdot x_i + b) \geq 1 - \xi_i \tag{11}$$

Point A

$$y_A(\vec{w}_A \cdot x_A + b) = \tag{12}$$

$$1(0 \cdot -5 + 2 \cdot 0 + -5) = -5 \tag{13}$$

Thus, $\xi_A = 6$

Computing slack

$$y_i(\vec{w}_i \cdot x_i + b) \geq 1 - \xi_i \tag{11}$$

Point A

$$y_A(\vec{w}_A \cdot x_A + b) = \tag{12}$$

$$1(0 \cdot -5 + 2 \cdot 0 + -5) = -5 \tag{13}$$

Thus, $\xi_A = 6$

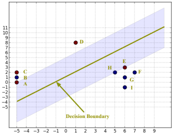
Point C

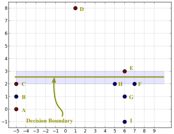
$$y_C(\vec{w}_C \cdot x_C + b) = \tag{14}$$

$$1(0 \cdot -5 + 2 \cdot 2 + -5) = -1 \tag{15}$$

Thus, $\xi_C = 2$

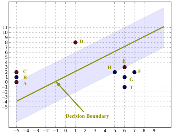
Which one is better?

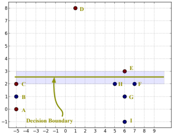




- Which decision boundary (wide / narrow) has the better objective?

Which one is better?

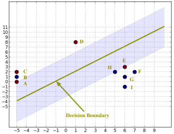


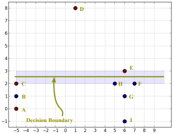


- Which decision boundary (wide / narrow) has the better objective?

$$\min_{w} \frac{1}{2}||w||^2 + C\sum_{i} \xi_i \tag{16}$$

Which one is better?





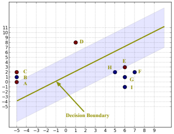
$$\frac{1}{2}||w||^2 = \frac{1}{2}\left(\frac{-1}{4}^2 + \frac{1}{4}^2\right) = 0.0625 \tag{16}$$

$$\sum_i \xi_i = 4.25 \tag{17}$$

- Which decision boundary (wide / narrow) has the better objective?

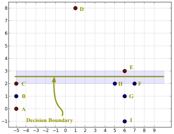
$$\min_w \frac{1}{2}||w||^2 + C\sum_i \xi_i \tag{18}$$

Which one is better?



$$\frac{1}{2}||w||^2 = 0.0625 \tag{16}$$

$$\sum_i \xi_i = 4.25 \tag{17}$$



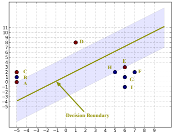
$$\frac{1}{2}||w||^2 = \frac{1}{2}\left(2^2\right) = 2 \tag{18}$$

$$\sum_i \xi_i = 8 \tag{19}$$

- Which decision boundary (wide / narrow) has the better objective?

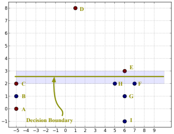
$$\min_w \frac{1}{2}||w||^2 + C\sum_i \xi_i \tag{20}$$

Which one is better?



$$\frac{1}{2}||w||^2 = 0.0625 \tag{16}$$

$$\sum_i \xi_i = 4.25 \tag{17}$$



$$\frac{1}{2}||w||^2 = 2 \tag{18}$$

$$\sum_i \xi_i = 8 \tag{19}$$

- Which decision boundary (wide / narrow) has the better objective?

$$\min_w \frac{1}{2}||w||^2 + C \sum_i \xi_i \tag{20}$$

- In this case it doesn't matter. Common C values: 1.0, $\frac{1}{m}$

Importance of C

- Need to do cross-validation to select C
- Don't trust default values
- Look at values with high ξ; are they bad data?

Importance of C

- Need to do cross-validation to select C
- Don't trust default values
- Look at values with high ξ; are they bad data?
- Next time: how to find w