Introduction to Machine Learning

Machine Learning: Jordan Boyd-Graber
University of Maryland

SLACK SVMS

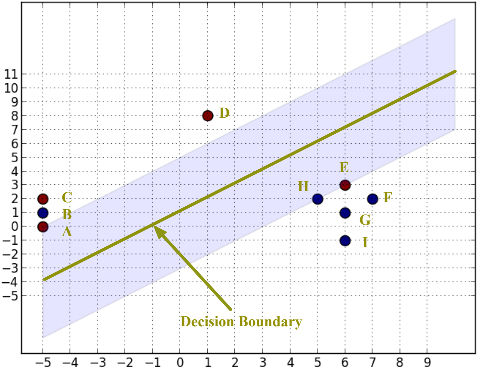
Administrivia Question

Content Question

Slack Example

Decision function:

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

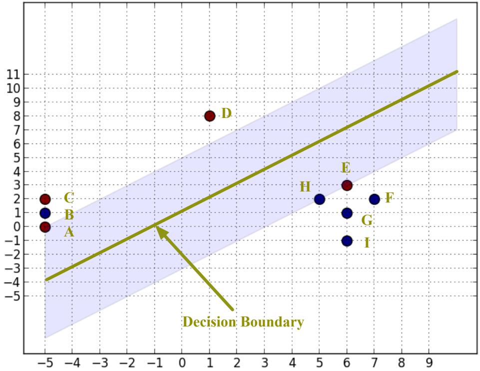


Slack Example

Decision function:

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

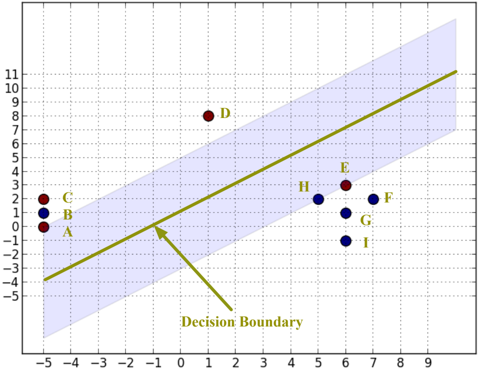
- What are the support vectors?



Slack Example

Decision function:

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



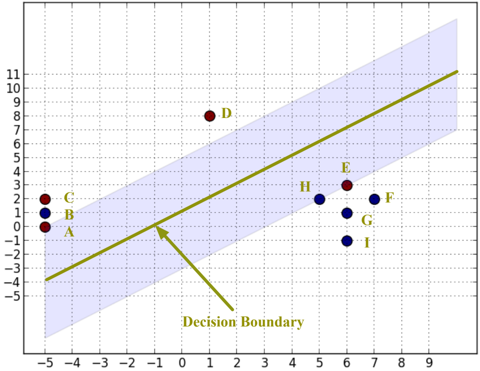
- What are the support vectors?
- Which have non-zero slack?

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

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

Computing slack

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

Point B

$$y_B(\vec{w}_B \cdot x_B + b) = \quad (2)$$

$$-1(-0.25 \cdot -5 + 0.25 \cdot 1 - 0.25) = -1.25 \quad (3)$$

Thus, $\xi_B = 2.25$

Computing slack

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

Point B

$$y_B(\vec{w}_B \cdot x_B + b) = \tag{2}$$

$$-1(-0.25 \cdot -5 + 0.25 \cdot 1 - 0.25) = -1.25 \tag{3}$$

Thus, $\xi_B = 2.25$

Point E

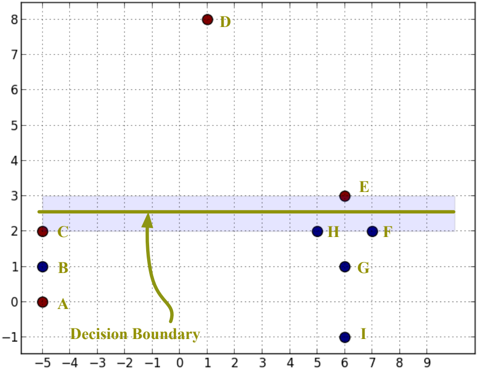
$$y_E(\vec{w}_E \cdot x_E + b) = \tag{4}$$

$$1(-0.25 \cdot 6 + 0.25 \cdot 3 + -0.25) = -1 \tag{5}$$

Slack Example

Decision function:

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

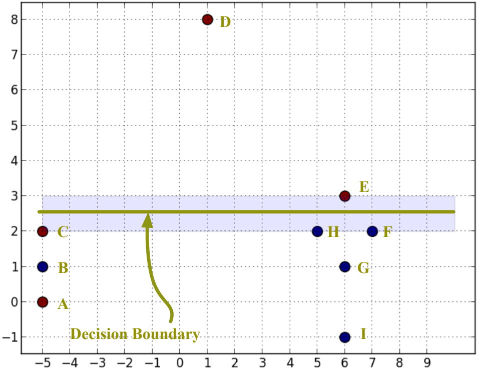


Slack Example

Decision function:

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

- What are the support vectors?

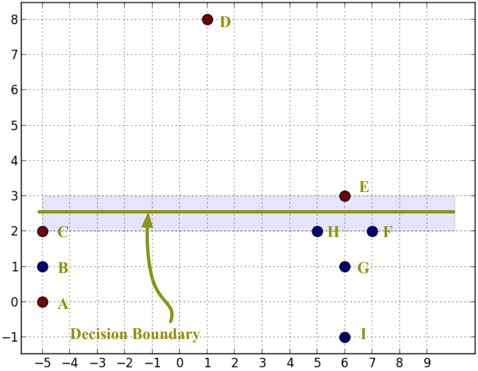


Slack Example

Decision function:

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

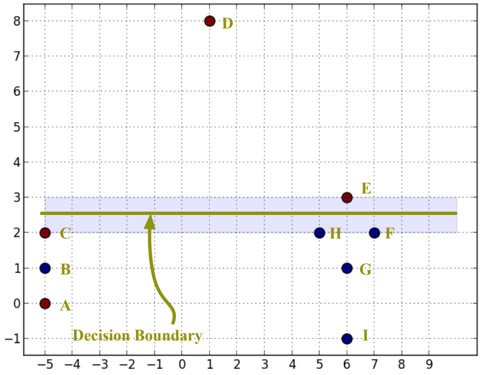
- What are the support vectors?
- Which have non-zero slack?



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

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

Computing slack

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

Point A

$$y_A(\vec{w}_A \cdot x_A + b) = \tag{7}$$

$$1(0 \cdot -5 + 2 \cdot 0 + -5) = -5 \tag{8}$$

Thus, $\xi_A = 6$

Computing slack

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

Point A

$$y_A(\vec{w}_A \cdot x_A + b) = \tag{7}$$

$$1(0 \cdot -5 + 2 \cdot 0 + -5) = -5 \tag{8}$$

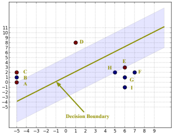
Thus, $\xi_A = 6$

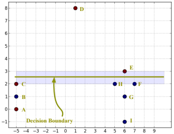
Point C

$$y_C(\vec{w}_C \cdot x_C + b) = \tag{9}$$

$$1(0 \cdot -5 + 2 \cdot 2 + -5) = -1 \tag{10}$$

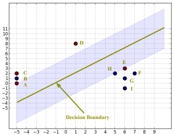
Which one is better?

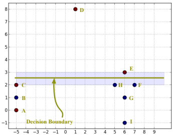




- Which decision boundary (wide / narrow) has the better objective?

Which one is better?

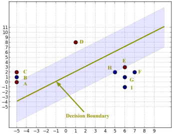




- Which decision boundary (wide / narrow) has the better objective?

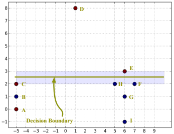
$$\min_{w} \frac{1}{2}||w||^2 + C\sum_{i} \xi_i \tag{11}$$

Which one is better?



$$\frac{1}{2}||w||^2 = \frac{1}{2}\left(\frac{-1}{4}^2 + \frac{1}{4}^2\right) = 0.0625 \tag{11}$$

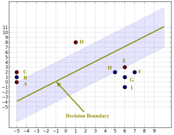
$$\sum_i \xi_i = 4.25 \tag{12}$$



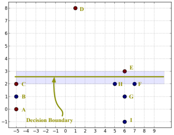
- Which decision boundary (wide / narrow) has the better objective?

$$\min_w \frac{1}{2}||w||^2 + C\sum_i \xi_i \tag{13}$$

Which one is better?



$$\frac{1}{2}||w||^2 = 0.0625 \tag{11}$$
$$\sum_i \xi_i = 4.25 \tag{12}$$

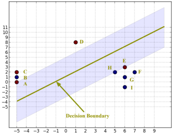


$$\frac{1}{2}||w||^2 = \frac{1}{2}\left(2^2\right) = 2 \tag{13}$$
$$\sum_i \xi_i = 8 \tag{14}$$

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

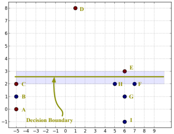
$$\min_w \frac{1}{2}||w||^2 + C\sum_i \xi_i \tag{15}$$

Which one is better?



$$\frac{1}{2}||w||^2 = 0.0625 \tag{11}$$

$$\sum_i \xi_i = 4.25 \tag{12}$$



$$\frac{1}{2}||w||^2 = 2 \tag{13}$$

$$\sum_i \xi_i = 8 \tag{14}$$

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

$$\min_w \frac{1}{2}||w||^2 + C \sum_i \xi_i \tag{15}$$

- 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?