Slack SVMs

Jordan Boyd-Graber
University of Colorado Boulder

LECTURE 8
Content Question
Content Question
Content Question
Content Question
Content Question
Content Question
Content Question
Content Question
Content Question
Content Question
Administrivia 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?
- Which have non-zero slack?
- Compute \( \xi_B, \xi_E \)
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?
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 \( \xi_B, \xi_E \)
Computing slack

\[ y_i (\vec{\omega}_i \cdot x_i + b) \geq 1 - \xi_i \]  \hspace{1cm} (1)
Computing slack

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

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

\[ -1(-0.25 \cdot -5 + 0.25 \cdot 1 - 0.25) = -1.25 \]

Thus, \( \xi_B = 2.25 \)
Computing slack

\[ y_i(\vec{w}_i \cdot x_i + b) \geq 1 - \xi_i \]  \hspace{1cm} (1)

Point B

\[ y_B(\vec{w}_B \cdot x_B + b) = \quad \hspace{1cm} (2) \]
\[ -1(-0.25 \cdot -5 + 0.25 \cdot 1 - 0.25) = -1.25 \]  \hspace{1cm} (3)

Thus, \( \xi_B = 2.25 \)

Point E

\[ y_E(\vec{w}_E \cdot x_E + b) = \quad \hspace{1cm} (4) \]
\[ 1(-0.25 \cdot 6 + 0.25 \cdot 3 + -0.25) = -1 \]  \hspace{1cm} (5)

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?
- Which have non-zero slack?
- Compute \( \xi_A \), \( \xi_C \)
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 \( \xi_A, \xi_C \)
Computing slack

\[ y_i(\vec{w}_i \cdot x_i + b) \geq 1 - \xi_i \]  \hspace{1cm} (6)
Computing slack

\[ y_i(\vec{w}_i \cdot x_i + b) \geq 1 - \xi_i \]  \hspace{1cm} (6)

**Point A**

\[ y_A(\vec{w}_A \cdot x_A + b) = \]
\[ 1(0 \cdot -5 + 2 \cdot 0 + -5) = -5 \]  \hspace{1cm} (8)

Thus, \( \xi_A = 6 \)
Computing slack

\[ y_i(\vec{w}_i \cdot x_i + b) \geq 1 - \xi_i \]  \hfill (6)

**Point A**

\[
y_A(\vec{w}_A \cdot x_A + b) =
\]
\[
1(0 \cdot -5 + 2 \cdot 0 + -5) = -5
\hfill (8)
\]

Thus, \( \xi_A = 6 \)

**Point C**

\[
y_C(\vec{w}_C \cdot x_C + b) =
\]
\[
1(0 \cdot -5 + 2 \cdot 2 + -5) = -1
\hfill (10)
\]

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?

\[
\min_w \frac{1}{2} ||w||^2 + C \sum_i \xi_i
\]  
(11)
Which one is better?

\[
\frac{1}{2} \left\| w \right\|^2 = \frac{1}{2} \left( \frac{-1^2}{4} + \frac{1^2}{4} \right) = 0.0625
\]

(11)

\[
\sum_i \xi_i = 4.25
\]

(12)

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

\[
\min_w \frac{1}{2} \left\| w \right\|^2 + C \sum_i \xi_i
\]

(13)
Which one is better?

\[
\frac{1}{2} \|w\|^2 = 0.0625 \quad (11) \quad \frac{1}{2} \|w\|^2 = \frac{1}{2} (2^2) = 2 \quad (13)
\]

\[
\sum_i \xi_i = 4.25 \quad (12) \quad \sum_i \xi_i = 8 \quad (14)
\]

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

\[
\min_w \frac{1}{2} \|w\|^2 + C \sum_i \xi_i \quad (15)
\]
Which one is better?

\[ \frac{1}{2} \|w\|^2 = 0.0625 \quad (11) \]
\[ \sum_i \xi_i = 4.25 \quad (12) \]
\[ \frac{1}{2} \|w\|^2 = 2 \quad (13) \]
\[ \sum_i \xi_i = 8 \quad (14) \]

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

\[ \min_w \frac{1}{2} \|w\|^2 + C \sum_i \xi_i \quad (15) \]

- 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 $\xi$; 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 $\xi$; are they bad data?
- Next time: how to find $w$