SVM

Introduction to Data Science Algorithms
Jordan Boyd-Graber and Michael Paul
SLIDES ADAPTED FROM JERRY ZHU
Can SVMs Work Here?

\[ y_i (w \cdot x_i + b) \geq 1 \]
Can SVMs Work Here?

\[ y_i (w \cdot x_i + b) \geq 1 \] (1)
Trick: Allow for a few bad apples
New objective function

\[
\min_{w, b, \xi} \frac{1}{2}||w||^2 + C \sum_{i=1}^{\xi} \xi_i^p
\]

subject to \( y_i(w \cdot x_i + b) \geq 1 - \xi_i \wedge \xi_i \geq 0, i \in [1, m] \)
New objective function

\[
\min_{w,b,\xi} \frac{1}{2}||w||^2 + C \sum_{i=1}^{\xi_i^p} \xi^p
\]

subject to \(y_i(w \cdot x_i + b) \geq 1 - \xi_i \land \xi_i \geq 0, i \in [1, m]\)

• Standard margin
New objective function

\[
\min_{w,b,\xi} \frac{1}{2} ||w||^2 + C \sum_{i=1}^{m} \xi_i^p
\]

subject to \(y_i(w \cdot x_i + b) \geq 1 - \xi_i \land \xi_i \geq 0, \ i \in [1, m]\)

- Standard margin
- How wrong a point is (slack variables)
New objective function

\[
\min_{w,b,\xi} \frac{1}{2}||w||^2 + C \sum_{i=1}^{\xi_i^p}
\]

subject to \(y_i (w \cdot x_i + b) \geq 1 - \xi_i \land \xi_i \geq 0, i \in [1, m]\)

- Standard margin
- How wrong a point is (slack variables)
- Tradeoff between margin and slack variables
New objective function

\[
\min_{w, b, \xi} \frac{1}{2}||w||^2 + C \sum_{i=1}^{m} \xi_i^p \\
\text{subject to } y_i(w \cdot x_i + b) \geq 1 - \xi_i \wedge \xi_i \geq 0, i \in [1, m]
\]

- Standard margin
- How wrong a point is (slack variables)
- Tradeoff between margin and slack variables
- How bad wrongness scales
Aside: Loss Functions

- Losses measure how bad a mistake is
- Important for slack as well
Aside: Loss Functions

- Losses measure how bad a mistake is
- Important for slack as well

Diagram showing the 0/1 Loss function.
 Aside: Loss Functions

- Losses measure how bad a mistake is
- Important for slack as well
Aside: Loss Functions

- Losses measure how bad a mistake is
- Important for slack as well
Aside: Loss Functions

- Losses measure how bad a mistake is
- Important for slack as well

We’ll focus on linear hinge loss

![Diagram showing different loss functions: Quadratic Hinge, Linear Hinge, and 0/1 Loss. The focus is on the linear hinge loss.](image)
Wrapup

- Adding slack variables don’t break the SVM problem
- Very popular algorithm
  - SVMLight (many options)
  - Libsvm / Liblinear (very fast)
  - Weka (friendly)
  - pyml (Python focused, from Colorado)