Optimizing Support Vector Machines

Jordan Boyd-Graber
University of Colorado Boulder
LECTURE 10

Slides adapted from David Page

Content Questions

Content Questions

Content Questions

Content Questions

Content Questions

Content Questions

Content Questions

Content Questions

Content Questions

Content Questions

Administrivia

- SVM homework and Boosting homework's posted
- Dates moved a week later for both

SMO Algorithm

Positive

(-2, 2)
(0, 4)
(2, 1)

Negative

(-2, -3)
(0, -1)
(2, -3)

SMO Algorithm

Positive	Negative
(-2, 2)	(-2, -3)
(0, 4)	(0, -1)
(2, 1)	(2, -3)

- Initially, all alphas are zero

$$\vec{\alpha} = < 0, 0, 0, 0, 0, 0 > \tag{1}$$

SMO Algorithm

Positive	Negative
(-2, 2)	(-2, -3)
(0, 4)	(0, -1)
(2, 1)	(2, -3)

- Initially, all alphas are zero

$$\vec{\alpha} = < 0, 0, 0, 0, 0, 0 > \quad (1)$$

- Intercept b is also zero
- Regularization $C = \pi$

SMO Optimization for $i = 0, j = 4$: Predictions and Step

- Prediction: $f(x_0)$
- Prediction: $f(x_4)$
- Error: E_0
- Error: E_4
- Step η

(2)

SMO Optimization for $i = 0, j = 4$: Predictions and Step

- Prediction: $f(x_0) = 0$
- Prediction: $f(x_4)$
- Error: E_0
- Error: E_4
- Step η

$$ \tag{2} $$

SMO Optimization for $i = 0, j = 4$: Predictions and Step

- Prediction: $f(x_0) = 0$
- Prediction: $f(x_4) = 0$
- Error: E_0
- Error: E_4
- Step η

(2)

SMO Optimization for $i = 0, j = 4$: Predictions and Step

- Prediction: $f(x_0) = 0$
- Prediction: $f(x_4) = 0$
- Error: $E_0 = -1$
- Error: $E_4 = +1$
- Step η

(2)

SMO Optimization for $i = 0, j = 4$: Predictions and Step

- Prediction: $f(x_0) = 0$
- Prediction: $f(x_4) = 0$
- Error: $E_0 = -1$
- Error: $E_4 = +1$
- Step η

$$\eta = 2\langle x_0, x_4 \rangle - \langle x_0, x_0 \rangle - \langle x_4, x_4 \rangle \tag{2}$$

SMO Optimization for $i = 0, j = 4$: Predictions and Step

- Prediction: $f(x_0) = 0$
- Prediction: $f(x_4) = 0$
- Error: $E_0 = -1$
- Error: $E_4 = +1$
- Step η

$$\eta = 2\langle x_0, x_4 \rangle - \langle x_0, x_0 \rangle - \langle x_4, x_4 \rangle = 2 \cdot -2 - 8 - 1 = -13 \quad (2)$$

SMO Optimization for $i = 0, j = 4$: Bounds

- Lower and upper bounds for α_j

$$L = \max(0, \alpha_j - \alpha_i) \tag{3}$$

$$H = \min(C, C + \alpha_j - \alpha_i) \tag{4}$$

SMO Optimization for $i = 0, j = 4$: Bounds

- Lower and upper bounds for α_j

$$L = \max(0, \alpha_j - \alpha_i) = 0 \tag{3}$$

$$H = \min(C, C + \alpha_j - \alpha_i) \tag{4}$$

SMO Optimization for $i = 0, j = 4$: Bounds

- Lower and upper bounds for α_j

$$L = \max(0, \alpha_j - \alpha_i) = 0 \tag{3}$$

$$H = \min(C, C + \alpha_j - \alpha_i) = \pi \tag{4}$$

SMO Optimization for $i = 0, j = 4$: α update

New value for α_j

$$\alpha_j^* = \alpha_j - \frac{y_j(E_i - E_j)}{\eta} \tag{5}$$

$$\tag{6}$$

SMO Optimization for $i = 0, j = 4$: α update

New value for α_j

$$\alpha_j^* = \alpha_j - \frac{y_j(E_i - E_j)}{\eta} = \frac{-2}{\eta} = \frac{2}{13} \tag{5}$$

$$\tag{6}$$

SMO Optimization for $i = 0, j = 4$: α update

New value for α_j

$$\alpha_j^* = \alpha_j - \frac{y_j(E_i - E_j)}{\eta} = \frac{-2}{\eta} = \frac{2}{13} \tag{5}$$

New value for α_i

$$\tag{6}$$

SMO Optimization for $i = 0, j = 4$: α update

New value for α_j

$$\alpha_j^* = \alpha_j - \frac{y_j(E_i - E_j)}{\eta} = \frac{-2}{\eta} = \frac{2}{13} \tag{5}$$

New value for α_i

$$\alpha_i^* = \alpha_i + y_i y_j \left(\alpha_j^{(old)} - \alpha_j\right) \tag{6}$$

SMO Optimization for $i = 0, j = 4$: α update

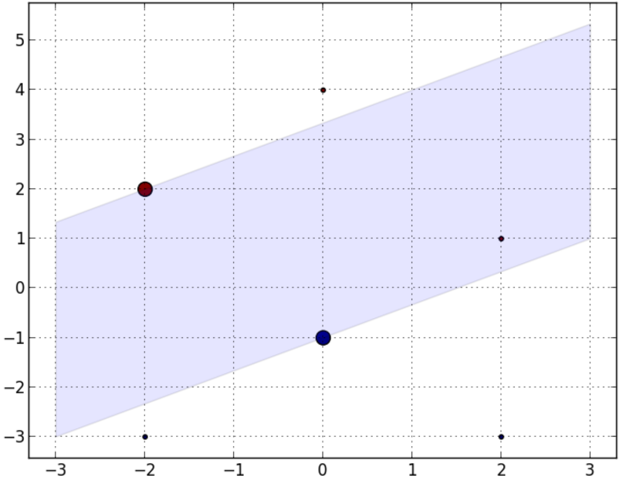
New value for α_j

$$\alpha_j^* = \alpha_j - \frac{y_j(E_i - E_j)}{\eta} = \frac{-2}{\eta} = \frac{2}{13} \tag{5}$$

New value for α_i

$$\alpha_i^* = \alpha_i + y_i y_j \left(\alpha_j^{(old)} - \alpha_j \right) = \alpha_j = \frac{2}{13} \tag{6}$$

Margin



Find weight vector and bias

- Weight vector

$$\vec{w} = \sum_{i=1}^{m} \alpha_i y_i \vec{x}_i \tag{7}$$

- Bias

$$b = b^{(old)} - E_i - y_i(\alpha_i^* - \alpha_i^{(old)}) x_i \cdot x_i - y_j(\alpha_j^* - \alpha_j^{(old)}) x_i \cdot x_j \tag{8}$$

$$\tag{9}$$

Find weight vector and bias

- Weight vector

$$\vec{w} = \sum_{i=1}^{m} \alpha_i y_i \vec{x}_i = \frac{2}{13}\begin{bmatrix} -2 \\ 2 \end{bmatrix} - \frac{2}{3}\begin{bmatrix} 0 \\ -1 \end{bmatrix} \quad (7)$$

- Bias

$$b = b^{(old)} - E_i - y_i(\alpha_i^* - \alpha_i^{(old)})x_i \cdot x_i - y_j(\alpha_j^* - \alpha_j^{(old)})x_i \cdot x_j \quad (8)$$

$$(9)$$

Find weight vector and bias

- Weight vector

$$\vec{w} = \sum_{i=1}^{m} \alpha_i y_i \vec{x}_i = \frac{2}{13} \begin{bmatrix} -2 \\ 2 \end{bmatrix} - \frac{2}{3} \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} \frac{-4}{13} \\ \frac{6}{13} \end{bmatrix} \quad (7)$$

- Bias

$$b = b^{(old)} - E_i - y_i(\alpha_i^* - \alpha_i^{(old)}) x_i \cdot x_i - y_j(\alpha_j^* - \alpha_j^{(old)}) x_i \cdot x_j \quad (8)$$

$$(9)$$

Find weight vector and bias

- Weight vector

$$\vec{w} = \sum_{i=1}^{m} \alpha_i y_i \vec{x}_i = \frac{2}{13} \begin{bmatrix} -2 \\ 2 \end{bmatrix} - \frac{2}{3} \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} \frac{-4}{13} \\ \frac{6}{13} \end{bmatrix} \tag{7}$$

- Bias

$$b = b^{(old)} - E_i - y_i(\alpha_i^* - \alpha_i^{(old)}) x_i \cdot x_i - y_j(\alpha_j^* - \alpha_j^{(old)}) x_i \cdot x_j \tag{8}$$

$$= 1 - \frac{2}{13} \cdot 8 + \frac{2}{13} \cdot -2 = -0.54 \tag{9}$$

SMO Optimization for $i = 2, j = 4$

Let's skip the boring stuff

- $E_2 = -1.69$
- $E_4 = 0.00$
- $\eta = -8$

- $\alpha_4 = \alpha_j^{(old)} - \frac{y_j(E_i - E_j)}{\eta}$
- $\alpha_2 = \alpha_i^{(old)} + y_i y_j \left(\alpha_j^{(old)} - \alpha_j \right)$

SMO Optimization for $i = 2, j = 4$

Let's skip the boring stuff

- $E_2 = -1.69$
- $E_4 = 0.00$
- $\eta = -8$

- $\alpha_4 = \alpha_j^{(old)} - \frac{y_j(E_i - E_j)}{\eta}$
- $\alpha_2 = \alpha_i^{(old)} + y_i y_j \left(\alpha_j^{(old)} - \alpha_j\right)$

SMO Optimization for $i = 2, j = 4$

Let's skip the boring stuff

- $E_2 = -1.69$
- $E_4 = 0.00$
- $\eta = -8$

- $\alpha_4 = \alpha_j^{(old)} - \frac{y_j(E_i - E_j)}{\eta} = 0.15 + \frac{-1.69}{-8} = 0.37$
- $\alpha_2 = \alpha_i^{(old)} + y_i y_j \left(\alpha_j^{(old)} - \alpha_j \right)$

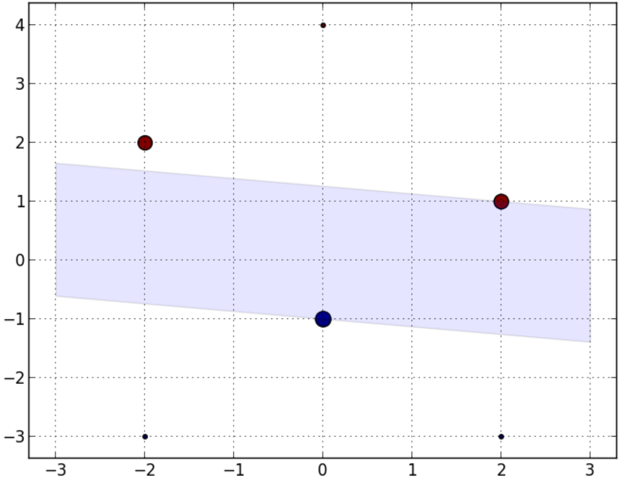
SMO Optimization for $i = 2, j = 4$

Let's skip the boring stuff

- $E_2 = -1.69$
- $E_4 = 0.00$
- $\eta = -8$

- $\alpha_4 = \alpha_j^{(old)} - \frac{y_j(E_i - E_j)}{\eta} = 0.15 + \frac{-1.69}{-8} = 0.37$
- $\alpha_2 = \alpha_i^{(old)} + y_i y_j \left(\alpha_j^{(old)} - \alpha_j\right) = 0 - (0.15 - 0.37) = 0.21$

Margin



Weight vector and bias

- Bias $b = -0.12$
- Weight vector

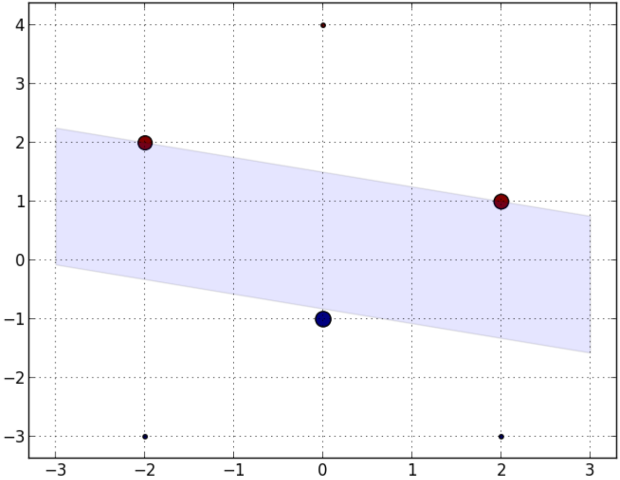
$$\vec{w} = \sum_{i=1}^{m} \alpha_i y_i \vec{x}_i \tag{10}$$

Weight vector and bias

- Bias $b = -0.12$
- Weight vector

$$\vec{w} = \sum_{i=1}^{m} \alpha_i y_i \vec{x}_i = \begin{bmatrix} 0.12 \\ 0.88 \end{bmatrix} \tag{10}$$

Another Iteration $(i = 0, j = 2)$



SMO Algorithm

- Convenient approach for solving: vanilla, slack, kernel approaches
- Convex problem
- Scalable to large datasets (implemented in scikit learn)
- What we didn't do:
  - Check KKT conditions
  - Randomly choose indices