



Inexact Search is “Good Enough”

Advanced Machine Learning for NLP

Jordan Boyd-Graber

MATHEMATICAL TREATMENT

Preliminaries: algorithm, separability

- Structured perceptron maintains set of “wrong features”

$$\Delta\vec{\Phi}(x, y, z) \equiv \vec{\Phi}(x, y) - \vec{\Phi}(x, z) \quad (1)$$

- Structured perceptron updates weights with

$$\vec{w} \leftarrow \vec{w} + \Delta\vec{\Phi}(x, y, z) \quad (2)$$

- Dataset D is linearly separable under features Φ with margin δ if

$$\vec{u} \cdot \Delta\vec{\Phi}(x, y, z) \geq \delta \quad \forall x, y, z \in D \quad (3)$$

Violations vs. Errors

- It may be difficult to find the highest scoring hypothesis
- It's okay as long as inference finds a **violation**

$$\vec{w} \cdot \Delta \vec{\Phi}(x, y, z) \leq 0 \quad (4)$$

- This means that y might not be answer algorithm gives

Limited number of mistakes

- Define diameter R as

$$R = \max_{(x,y,z)} \|\Delta\vec{\Phi}(x,y,z)\| \quad (5)$$

Limited number of mistakes

- Define diameter R as

$$R = \max_{(x,y,z)} \|\Delta\vec{\Phi}(x,y,z)\| \quad (5)$$

- Weight vector \vec{w} grows with each error
- We can prove that $\|\vec{w}\|$ can't get too big
- And thus, algorithm can only run for limited number of iterations k where it updates weights
- Indeed, we'll bound it from two directions

$$k^2 \delta^2 \leq \|w^{(k+1)}\|^2 \leq kR^2 \quad (6)$$

Lower Bound

Lower Bound

$$k^2 \delta^2 \leq \|w^{(k+1)}\|^2$$

(7)

Lower Bound

Lower Bound

$$k^2 \delta^2 \leq \|w^{(k+1)}\|^2$$

$$\vec{w}^{(k+1)} = w^{(k)} + \Delta \vec{\Phi}(x, y, z) \tag{7}$$

(8)

Update equation

Lower Bound

Lower Bound

$$k^2 \delta^2 \leq \|w^{(k+1)}\|^2$$

$$\vec{w}^{(k+1)} = w^{(k)} + \Delta \vec{\Phi}(x, y, z) \quad (7)$$

$$\vec{u} \cdot \vec{w}^{(k+1)} = \vec{u} \cdot w^{(k)} + \vec{u} \Delta \vec{\Phi}(x, y, z) \quad (8)$$

$$(9)$$

Multiply both sides by \vec{u}

Lower Bound

Lower Bound

$$k^2 \delta^2 \leq \|w^{(k+1)}\|^2$$

$$\vec{w}^{(k+1)} = w^{(k)} + \Delta \vec{\Phi}(x, y, z) \quad (7)$$

$$\vec{u} \cdot \vec{w}^{(k+1)} = \vec{u} \cdot w^{(k)} + \vec{u} \Delta \vec{\Phi}(x, y, z) \quad (8)$$

$$\vec{u} \cdot \vec{w}^{(k+1)} \geq \vec{u} \cdot w^{(k)} + \delta \quad (9)$$

Definition of margin

Lower Bound

Lower Bound

$$k^2 \delta^2 \leq \|w^{(k+1)}\|^2$$

$$\vec{w}^{(k+1)} = w^{(k)} + \Delta \vec{\Phi}(x, y, z) \quad (7)$$

$$\vec{u} \cdot \vec{w}^{(k+1)} = \vec{u} \cdot w^{(k)} + \vec{u} \Delta \vec{\Phi}(x, y, z) \quad (8)$$

$$\vec{u} \cdot \vec{w}^{(k+1)} \geq \vec{u} \cdot w^{(k)} + \delta \quad (9)$$

By induction, $\vec{u} \cdot \vec{w}^{(k+1)} \geq k\delta$ (Base case: $\vec{w}^0 = \vec{0}$)

Lower Bound

Lower Bound

$$k^2 \delta^2 \leq \|w^{(k+1)}\|^2$$

$$\vec{u} \cdot \vec{w}^{(k+1)} \geq \vec{u} \cdot w^{(k)} + \delta \quad (7)$$

By induction, $\vec{u} \cdot \vec{w}^{(k+1)} \geq k\delta$ (Base case: $\vec{w}^0 = \vec{0}$)

$$\|\vec{u}\| \|\vec{w}^{(k+1)}\| \geq \vec{u} \cdot \vec{v} \geq k\delta \quad (8)$$

For any vectors, $\|\vec{a}\| \|\vec{b}\| \geq a \cdot b$

Lower Bound

Lower Bound

$$k^2 \delta^2 \leq \|w^{(k+1)}\|^2$$

$$\vec{u} \cdot \vec{w}^{(k+1)} \geq \vec{u} \cdot w^{(k)} + \delta \quad (7)$$

By induction, $\vec{u} \cdot \vec{w}^{(k+1)} \geq k\delta$ (Base case: $\vec{w}^0 = \vec{0}$)

$$\|\vec{u}\| \|\vec{w}^{(k+1)}\| \geq \vec{u} \cdot \vec{w} \geq k\delta \quad (8)$$

$$\|\vec{w}^{(k+1)}\| \geq k\delta \quad (9)$$

\vec{u} is a unit vector

Lower Bound

Lower Bound

$$k^2 \delta^2 \leq \|w^{(k+1)}\|^2$$

$$\vec{u} \cdot \vec{w}^{(k+1)} \geq \vec{u} \cdot w^{(k)} + \delta \quad (7)$$

By induction, $\vec{u} \cdot \vec{w}^{(k+1)} \geq k\delta$ (Base case: $\vec{w}^0 = \vec{0}$)

$$\|\vec{u}\| \|\vec{w}^{(k+1)}\| \geq \vec{u} \cdot \vec{v} \geq k\delta \quad (8)$$

$$\|\vec{w}^{(k+1)}\| \geq k\delta \quad (9)$$

$$\|\vec{w}^{(k+1)}\|^2 \geq k^2 \delta^2 \quad (10)$$

Square both sides, and we're done!

Upper Bound

Upper Bound

$$\|\vec{w}^{(k+1)}\|^2 \leq kR^2 \quad (11)$$

(12)

Upper Bound

Upper Bound

$$\|\vec{w}^{(k+1)}\|^2 \leq kR^2 \quad (11)$$

$$\|\vec{w}^{(k+1)}\|^2 = \|\vec{w}^{(k)} + \Delta\vec{\Phi}(x, y, z)\|^2 \quad (12)$$

Update rule

Upper Bound

Upper Bound

$$\|\vec{w}^{(k+1)}\|^2 \leq kR^2 \quad (11)$$

$$\|\vec{w}^{(k+1)}\|^2 = \|\vec{w}^{(k)} + \Delta\vec{\Phi}(x, y, z)\|^2 \quad (12)$$

$$\|\vec{w}^{(k+1)}\|^2 = \|\vec{w}^{(k)}\|^2 + \|\Delta\vec{\Phi}(x, y, z)\|^2 + 2\vec{w}^{(k)} \cdot \Delta\vec{\Phi}(x, y, z) \quad (13)$$

Law of cosines

Upper Bound

Upper Bound

$$\|\vec{w}^{(k+1)}\|^2 \leq kR^2 \quad (11)$$

$$\|\vec{w}^{(k+1)}\|^2 = \|\vec{w}^{(k)} + \Delta\vec{\Phi}(x, y, z)\|^2 \quad (12)$$

$$\|\vec{w}^{(k+1)}\|^2 = \|\vec{w}^{(k)}\|^2 + \|\Delta\vec{\Phi}(x, y, z)\|^2 + 2w^{(k)} \cdot \Delta\vec{\Phi}(x, y, z) \quad (13)$$

$$\|\vec{w}^{(k+1)}\|^2 \leq \|\vec{w}^{(k)}\|^2 + R^2 + 2w^{(k)} \cdot \Delta\vec{\Phi}(x, y, z) \quad (14)$$

Definition of diameter

Upper Bound

Upper Bound

$$\|\vec{w}^{(k+1)}\|^2 \leq kR^2 \quad (11)$$

$$\|\vec{w}^{(k+1)}\|^2 = \|\vec{w}^{(k)} + \Delta\vec{\Phi}(x, y, z)\|^2 \quad (12)$$

$$\|\vec{w}^{(k+1)}\|^2 = \|\vec{w}^{(k)}\|^2 + \|\Delta\vec{\Phi}(x, y, z)\|^2 + 2\vec{w}^{(k)} \cdot \Delta\vec{\Phi}(x, y, z) \quad (13)$$

$$\|\vec{w}^{(k+1)}\|^2 \leq \|\vec{w}^{(k)}\|^2 + R^2 + 2\vec{w}^{(k)} \cdot \Delta\vec{\Phi}(x, y, z) \quad (14)$$

$$\|\vec{w}^{(k+1)}\|^2 \leq \|\vec{w}^{(k)}\|^2 + R^2 + 0 \quad (15)$$

If violation

Putting it together

- Sandwich:

$$k^2 \delta^2 \leq \|w^{(k+1)}\|^2 \leq kR^2 \quad (16)$$

Putting it together

- Sandwich:

$$k^2 \delta^2 \leq \|w^{(k+1)}\|^2 \leq kR^2 \quad (16)$$

- Solve for k :

$$k \leq \frac{R^2}{\delta^2} \quad (17)$$

Putting it together

- Sandwich:

$$k^2 \delta^2 \leq \|w^{(k+1)}\|^2 \leq kR^2 \quad (16)$$

- Solve for k :

$$k \leq \frac{R^2}{\delta^2} \quad (17)$$

- What does this mean?

Putting it together

- Sandwich:

$$k^2 \delta^2 \leq \|w^{(k+1)}\|^2 \leq kR^2 \quad (16)$$

- Solve for k :

$$k \leq \frac{R^2}{\delta^2} \quad (17)$$

- What does this mean?
- Limited number of errors (updates)
 - Larger diameter increases errors (worst possible mistake)
 - Larger margin decreases errors (bigger separation from wrong answer)
- Finding the largest violation wrong answer is best (but any violation okay)

In Practice

Harder the search space, the more max violation helps

