Support Vector Machines

SUE ANN HONG
10/18/2007

THE MOST FAMOUS SLIDE *EVER*
Linear classifier: $h(x) = \text{sign}(w.x + b)$

- $|w.x+b|$ big if far away from the boundary "confidence"
PRIMAL: THE “INTUITIVE” VERSION

$$\min \|w\|^2 + c \sum \xi$$

s.t. $$(w.x + b)y \geq 1 - \xi$$

$$\xi \geq 0$$

- (w.x+b)y > 0  iff  w.x+b and y same sign
- so want (w.x+b)y to be as large as possible
- could set \(\xi\)’s to anything big... no constraint?!
- need \(\xi\)’s to be small... but set \(\xi < 0\) for a confident point, \(\xi\) can be big for some other point
- \(\xi \geq 0\) means no love shared

1. regularizer
2. \(-1/margin\)

How do we find \(w\)?

\(\text{Quadratic programming}\)

How do we find \(C\)?

\(\text{Cross-validation!}\)

HW3! :)

get your libsvm today!
**DUAL: THE “SUPPORT VECTOR” VERSION**

\[
\begin{align*}
\max & \sum \alpha - \frac{1}{2} \sum \alpha_i \alpha_j y_i y_j x_i x_j \\
\text{s.t.} & \sum \alpha_i y_i = 0 \\
& c \geq \alpha_i \geq 0
\end{align*}
\]

- Where did this come from?
- Remember Lagrange Multipliers
  - Let us “incorporate” constraints into objective
  - Then solve the problem in the “dual” space of lagrange multipliers

**PRIMAL FEAR**

\[
\begin{align*}
\min & ||w||^2 + c \sum \xi \\
\text{s.t.} & (w.x + b) y \geq 1 - \xi \\
& \xi \geq 0
\end{align*}
\]

\[
\begin{align*}
\max & \sum \alpha - \frac{1}{2} \sum \alpha_i \alpha_j y_i y_j x_i x_j \\
\text{s.t.} & \sum \alpha_i y_i = 0 \\
& c \geq \alpha_i \geq 0
\end{align*}
\]

- Number of parameters?
  - large # features?
  - large # examples?
    - for large # features, DUAL preferred
    - many \( \alpha_i \) can go to zero!
DUAL: THE “SUPPORT VECTOR” VERSION

\[ \max \sum \alpha - \frac{1}{2} \sum \alpha_i \alpha_j y_i y_j x_i x_j \]
\[ \text{s.t. } \sum \alpha_i y_i = 0 \]
\[ C \geq \alpha_i \geq 0 \]

\[ \square \text{ Wait... how do we predict } y \]
\[ \text{for a new point } x?\]
\[ y = \text{sign}(w.x + b) \]

\[ \square \text{ How do we find } w? \]
\[ w = \sum \alpha_i y_i x_i \]

\[ \square \text{ How do we find } C? \]
\[ \text{Cross-validation!} \]

\[ \square \text{ b? “intersection” (algebra I),} \]
\[ y = \text{sign}(\sum \alpha_i y_i x_i x_j + b) \]

“SUPPORT VECTOR”S?

\[ \max \sum \alpha - \frac{1}{2} \sum \alpha_i \alpha_j (1)(1) \]
\[ \text{s.t. } \sum \alpha_i y_i = 0 \]
\[ C \geq \alpha_i \geq 0 \]

\[ \max \sum \alpha - \alpha_1 \alpha_2 (-1)(0+2) \]
\[ -1/2 \alpha_1^2 (1)(0+1) \]
\[ -1/2 \alpha_2^2 (1)(4+4) \]

\[ \max \alpha_1 + \alpha_2 + 2\alpha_1 \alpha_2 - \alpha_1^2 / 2 - 4\alpha_2^2 \]
\[ \text{s.t. } \alpha_1 - \alpha_2 = 0 \]
\[ C \geq \alpha_i \geq 0 \]

\[ \alpha_1 = \alpha_2 = \alpha \]
\[ \max 2\alpha - 5/2\alpha^2 \]
\[ \max 5/2\alpha(4/5-\alpha) \]

\[ b \]
\[ y = w.x + b \]
\[ b = y - w.x \]
\[ x_1: b = 1 - .4 [-2 -1][0 1] \]
\[ = 1 + .4 = 1.4 \]

\[ \alpha_1 = \alpha_2 = 2/5 \]
\[ w = \sum \alpha_i y_i x_i \]
\[ w = .4[0 1][-2 2] \]
\[ = .4[-2 -1] \]
"SUPPORT VECTOR"S?

\[
\begin{align*}
\max & \sum \alpha - \frac{1}{2} \sum \alpha_i \alpha_j y_i y_j x_i x_j \\
\text{s.t.} & \sum \alpha_i y_i = 0 \\
& C \geq \alpha_i \geq 0
\end{align*}
\]

\[
\max \alpha_1 + \alpha_2 + \alpha_3 + 2\alpha_1 \alpha_2 + 2\alpha_2 \alpha_3 - \alpha_1^2/2 - 4\alpha_2^2 - \alpha_3^2/2 \\
\text{s.t.} \alpha_1 - \alpha_2 + \alpha_3 = 0
\]

\[
\alpha_3 = \alpha_1 + \alpha_3 \quad \text{let } \alpha_2 = k \\
\alpha_1 + \alpha_3 = k
\]

\[
\begin{align*}
\max & \left( \alpha_1 + \alpha_3 \right) + \alpha_2 + 2\alpha_1 \alpha_3 \left( \alpha_1 + \alpha_3 \right) - \alpha_1^2/2 - 4\alpha_2^2 - \alpha_3^2/2 \\
\max & \left( \alpha_1^2 - \alpha_3^2 \right)/2 \\
\max & \left( \alpha_1^2 - \left( k - \alpha_1 \right)^2 \right)
\end{align*}
\]

What is \( \alpha_5 \)?
Try this at home

\[
b = y - w.x...
\]
Which ones are support vectors?
Why?
intuition: how many points “define” a line in 2D?
HW3
**“SUPPORT VECTOR”’S?**

\[
\begin{align*}
&\max \sum \alpha - \frac{1}{2} \sum \alpha_i \alpha_j y_i y_j x_i x_j \\
&\text{s.t. } \sum \alpha_i y_i = 0 \\
&\quad c \geq \alpha_i \geq 0
\end{align*}
\]

\[
\text{Loss part: } c \sum \xi \quad ||w||^2 \text{ ~regularization}
\]

1. \( \xi \geq 0 \) only if \( (w.x + b)y < 1 \)
2. we want: \( \xi \geq 1 - (w.x + b)y \) \( \xi \) minimize \( \xi \)

\[\Rightarrow \xi = 1 - (w.x + b)y\]

\[\Rightarrow \text{loss} = c(1 - (w.x + b)y) \text{ only if } (w.x + b)y < 1\]

**HINGE LOSS YOUR LOSS**
Hinge loss

\[ L = 1 - (w \cdot x + b)y \] only if \( (w \cdot x + b)y < 1 \)

**Hinge Loss Your Loss**

\[ z = (w \cdot x + b)y \]

**Loss for Positive Class**

\[ (w \cdot x + b) \]
Hinge Loss

1. Hinge Loss
   \[ L = 1 - (w \cdot x + b) y \quad \text{only if} \quad (w \cdot x + b) y < 1 \]

2. 0/1 Loss
   \[ L = 1 \quad \text{if} \quad (w \cdot x + b) y < 0, \quad 0 \quad \text{otherwise} \]
HINGE LOSS YOUR LOSS

- Hinge loss: \( L = 1 - (w \cdot x + b) y \) only if \((w \cdot x + b) y < 1\)
- 0/1 loss: \( L = 1 \) if \((w \cdot x + b) y < 0\), 0 otherwise
- Logistic loss: \( L = ||w||^2 + \sum \ln P(Y=1|x,w) = \ldots \ln(1+e^{-w \cdot x + b})\)

---

SVM

- Decision boundary: plain SVM is quite simple
- Why is the dual form useful? interesting?
- Support vectors are neat! (computationally, kernel trick, ...)
- SVM, LR, Boosting, ... all a family (diff loss)
- Which letters didn’t we see? C b \( \xi \) \( \alpha \) \( \gamma \) \( \mu \) x y z w t f?
LAST REMARKS

- I <3 Burges' tutorial -- READ IT!!!
  - first part (VC-dim, etc) might not make sense until learning theory lectures, but charge on

- MIDTERM REVIEW ON TUESDAY
  - 5-6:30, tentatively, somewhere