FUN SVM TRICKS

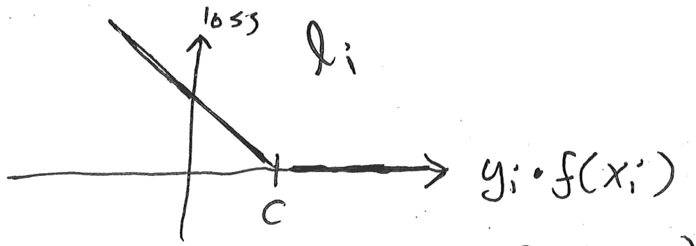
1. ONE COOL TRICK TO LEARN LINEAR SVM WEIGHTS WITHOUT SOLVING A QP.
2. ADAPTING SVMs TO LEARN RANKINGS
3. DROP OUT OF SCHOOL TO MAKE $$ WITH SUPPORT VECTOR RANKING

SVM - MARGIN CLASSIFICATION

WE WANT $y_i \cdot f(x_i) > c$ $\qquad x_i \in \mathbb{R}^n, \; y_i \in \{-1, +1\}$

FOR ALL $i = 1, \ldots, M$

HINGE LOSS = AMOUNT THIS INEQUALITY IS VIOLATED



(HAPPY WHEN $y_i \cdot f(x_i) > c$)

OVERALL LOSS: $\quad l = \left[\sum_{i=1}^{M} \max(0, \, c - y_i \cdot f(x_i)) \right] +$ regularization term

FOR LINEAR LEARNER, $\quad f(x_i) = w^T x_i = w \cdot x_i \quad$ FOR $w \in \mathbb{R}^n$ *

* BIAS/INTERCEPT: TACK A 1 ONTO EACH DATA POINT...

SO $$l(w) = \left[\sum_{i=1}^{M} \max(0, \, c - y_i \cdot w^T x_i) \right] + \frac{1}{2} \|w\|^2$$

$$= \left[\sum_{i=1}^{M} l_i(w) \right] + \frac{1}{2} \|w\|^2$$

WE COULD MINIMIZE $\ell(w)$ BY GRADIENT DESCENT IF WE KNEW ITS GRADIENT!

$$\nabla \ell(w) = \sum_i \nabla \ell_i(w) + w$$

$$\nabla \ell_i(w)^{*} = \frac{d}{dw} \max(0, c - y_i \cdot w^T x_i)$$

$$= \begin{cases} 0 \in \mathbb{R}^n & \text{if } y_i \cdot w^T x_i > c \\ -y_i x_i & \text{otherwise} \end{cases}$$

* TECHNICALLY THIS IS THE SUBGRADIENT B/C MAX NOT DIFFERENTIABLE.
IN PRACTICE: WORKS GREAT!

LETS USE STOCHASTIC GRADIENT DESCENT to DIRECTLY MINIMIZE $\ell(w)$:

"MINI-BATCH"

1. CHOOSE A RANDOM SUBSET OF TRAINING DATA
2. COMPUTE $\nabla \ell(w)$ ON THAT SUBSET
3. UPDATE $w \Leftarrow w - \alpha \nabla \ell(w)$ FOR A SMALL STEP SIZE $\alpha > 0$ *
4. REPEAT!

* THINGS CONVERGE NICELY IF YOU DECAY α OVER TIME...

LETS TRY IT IN CODE!

DISCLAIMER: ITS USUALLY BETTER TO USE PREEXISTING SVM SOLVERS!
LIBLINEAR & LIBSVM ARE GREAT AND THEY'RE BUILT INTO sklearn...

... BUT PRETTY COOL TO SEE WE CAN SOLVE SMALL PROBLEMS IN <100 LINES OF PYTHON!

SUPPORT VECTOR RANKING AKA ORDINAL REGRESSION

TRADITIONAL SVM FOR BINARY CLASSIFICATION INPUTS ARE $x_i \in \mathbb{R}^n$, $y_i \in \{-1, +1\}$

CONSTRAINT IS $y_i \cdot f(x_i) > c$

ORDINAL REGRESSION: INPUTS ARE x_i^+, x_i^- IN $\mathbb{R}^n$

CONSTRAINT IS $f(x_i^+) - f(x_i^-) > c$

FOR LINEAR SVM, $w^T(x_i^+ - x_i^-) > c \Leftrightarrow w^T \Delta x_i > c$ WHERE $\Delta x_i = x_i^+ - x_i^-$

f SCORES POSITIVE SAMPLE ABOVE NEGATIVE SAMPLE BY A MARGIN!

(LITTLEDOG SLIDES) A* FOR FOOTSTEP PLANNING THRU MOVIE & LEARNED COSTS x4

TAKING SVMS TO VEGAS

LET'S PREDICT NCAA OUTCOMES! [SHOW EXCEL]

FEATURE ENCODING

$$x = \begin{bmatrix} \text{team A} \\ \text{team B} \\ \text{team C} \\ \vdots \\ \text{home} \end{bmatrix} \quad \text{all 0 or 1}$$

SO IF TEAM A (HOME) BEATS TEAM B (AWAY)

$$x^+ = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \qquad x^- = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\Delta x = x^+ - x^- = \begin{bmatrix} 1 \\ -1 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

I WANT $w^T \Delta x > c$

NUDGE WEIGHT FOR TEAM A ↑

" " " B ↓

COMPARE TO REGRESSION LEARNER

$f(\Delta x) \mapsto$ POINT SPREAD

$w^T \Delta x$ = point differential

SURPRISINGLY, MARGIN LEARNER WORKS JUST AS WELL! [RUN DEMO]

IF YOU LIKE USING MATH + ML FOR SPORTS PREDICTION, CHECK OUT TIM CHARTIERS @ DAVIDSON! HE WAS THE ONE WHO ORIGINALLY INSPIRED ME TO SCOPE OUT MARGIN LEARNERS FOR SPORTS!

DISCLAIMER: STAY IN SCHOOL. INDIVIDUAL RESULTS MAY VARY