1

SVM

MENTION ESL
BIG PICTURE

(1)

OPT. $\underset{w, w_0}{\text{argmin}} \; \frac{1}{2}\|w\|$

s.t. $y_i(w^T x_i + w_0) \geq +1 \quad \forall i$

ρ is distance b/w points on +HP TO DIVIDING HP

$r = \frac{f(x)}{\|w\|} = \frac{1}{\|w\|}$

$\|w\| \rho = 1$ NOT CONSTRAINT

(2)

GEOW DUAL: $\underset{\alpha_i}{\text{argmax}} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \, x_i^T x_j$

$w = \sum_i \alpha_i y_i x_i$

LOTS OF VARIATIONS OF SVM
PLUG INTO QUADPROG AND EXPERIMENT.

1. SOFT MARGIN

$\underset{w, w_0, \varepsilon}{\text{argmin}} \; \frac{1}{2}\|w\|^2 + \sum_i^c \varepsilon_i^2$ s.t. $y_i(w^T x_i + w_0) \geq 1 - \varepsilon_i \quad \forall i$

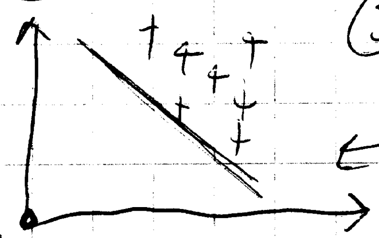
$\varepsilon_i \geq 0$

WHY? NEED NEAR L.S.
BUT IF MORE POINTS CROSS THEN BAD.

2. ONE CLASS SVM

EXAMPLES IN PROJECT — NO -VE's

1. ~~INTRUSION~~ OUTLIER DETECTION
2. RETRIEVAL PROBLEM → requires COUNT DATA
3. INLIER DETECTION



-ve

$$r = (w^T x + w_0) / \|w\|$$

For any support vector (instance the +ve or -ve hyperplane goes through

$$r = y_i (w^T x_i + w_0) / \|w\|$$

we want to maximize the margin

$$y_i (w^T x_i + w_0) / \|w\| \geq \rho \ \forall i$$

(1)

SVM

MENTION ESL

BIG PICTURE

(1)

OPT: $\underset{w, w_0}{\text{argmin}} \ \frac{1}{2}\|w\|$

s.t. $y_i(w^T x_i + w_0) \geq +1 \quad \forall i$

ρ is distance b/w points on + HP TO DIVIDING HP

$r = \frac{f(x)}{\|w\|} = \frac{1}{\|w\|}$

$\|w\| \rho = 1$ ← NOT CONSTRAINT

(2)

geow DUAL: $\underset{\alpha_i}{\text{argmax}} \sum_i \sum_j \alpha_i \alpha_j \, y_i y_j \, x_i^T x_j$

$w = \sum_i \alpha_i y_i x_i$

LOTS OF VARIATIONS OF SVM

PLUG INTO QUADPROG AND EXPERIMENT.

① SOFT MARGIN

$\underset{w, w_0, \varepsilon}{\text{argmin}} \ \frac{1}{2}\|w\|^2 + \sum_i \varepsilon_i^2$ s.t. $y_i(w^T x_i + w_0) \geq 1 - \varepsilon_i \quad \forall i$

$\varepsilon_i \geq 0$

WHY? NEED NEAR L.S.

BUT IF MORE POINTS CROSS THEN BAD.

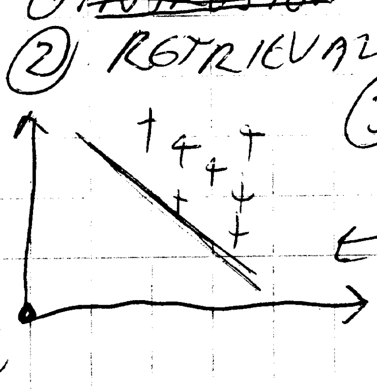
② ONE CLASS SVM

EXAMPLES IN PROJECT — NO -VE's

① ~~INTRUSION~~ OUTLIER DETECTION

② RETRIEVAL PROBLEM → requires COUNT DATA

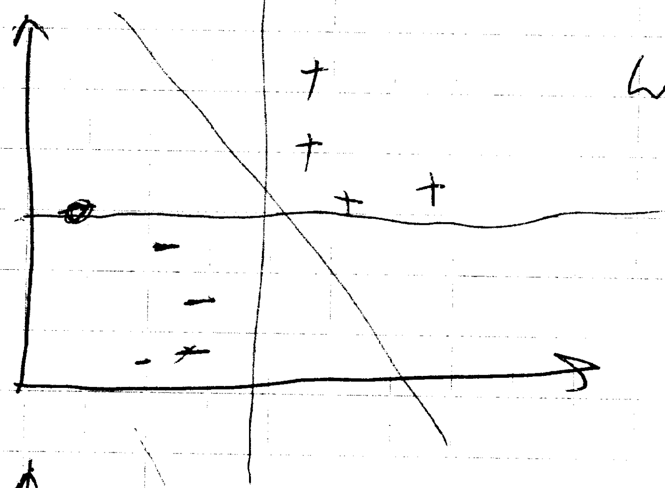
③ INLIER DETECTION



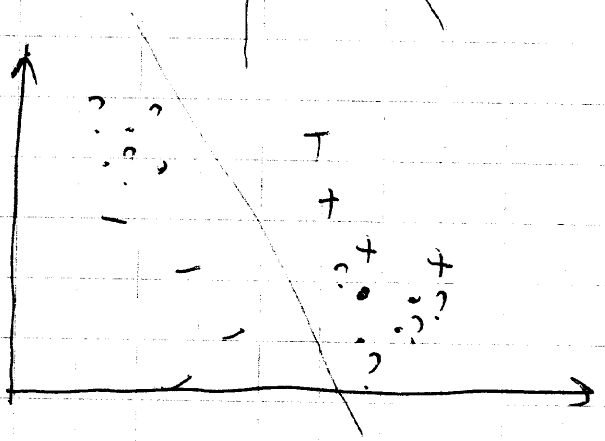
2

3 SEMI-SUPERVISED + TRANSDUCTION

1 SSL



WHICH ONE



WHY X UNLABELED?

DIFFICULT TO FIND LABELS OR LABEL

2 TRANSDUCTION

1 LOCATION STUDIES

2 PREDICTING GRADES IN THIS CLASS

3 PERSONAL INFO MANAGEMENT PROBLEMS

$$\min_{w, \hat{y}_1 \ldots \hat{y}_m} \frac{1}{2}\|w\|^2 \quad \text{s.t.} \quad y_i(w^T x_i + w_0) \geq 1 \ \forall i$$
$$\hat{y}_j(w^T x_j + w_0) \geq 1 \ \forall j$$

NO LONGER CONVEX. $\hat{y}_j \in \{-1, +1\} \ \forall j$

TRANSFER LEARNING

$X \rightarrow Y$"

SOURCE, TARGET

ELEPHANTS VS RHINO — SHEEP VS GOAT

ASSUMPTION SOURCE IS EASY TARGET "HARD"

HARD FOR MANY REASON: ① Few labels

② difficult problem

HOW CAN SOURCE TARGET DIFFER

$P_S(X_S) \neq P_T(X_T)$ $\quad P_S(Y|X) = P_T(Y|X)$

MEANS:

DOMAIN ADAPTATION — age

Can we adapt assume

$P_S(X) = P_T(X)$ $\quad P_S(Y|X) \neq P_T(Y|X)$

TASK TRANSFER

$$\underset{w, w_0}{\operatorname{argmin}} \ \tfrac{1}{2}\|w\|^2 + C \sum_{i=1}^{n} \xi_i^2$$

s.t. $\xi_i \geq 0 \ \forall i$

$y_i f^S(x_i) + y_i(w^T x_i + w_0) \geq 1 - \xi_i \quad \forall i$

$$f(x_i) = f^S(x_i) + w_i x_i + w_0$$

(4)

ALL OF THIS ASSUMES L.S. OR CLOSE (ENOUGH) SOL^{ns}:

$$\underset{\alpha_i}{\operatorname{argmax}} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \overbrace{x_i^T x_j}^{I.P}$$

$$X_i = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad X_j = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$X_i^T x_j = [4]$ LINEAR KERNEL

$(X_i X_j)^2 = 16$ POLYNOMIAL KERNEL

$$\left([X_{11} \; X_{12}] \begin{bmatrix} X_{21} \\ X_{22} \end{bmatrix} \right)^2$$

$$= (X_{11} X_{21} + X_{12} X_{22})^2$$

$$= (X_{11} X_{21})^2 + 2 X_{11} X_{21} X_{12} X_{22} + (X_{12} X_{22})^2$$

$$= (X_{11})^2 (X_{21})^2 + \sqrt{2} X_{11} X_{12} \sqrt{2} X_{21} X_{22} + (X_{12})^2 \times (X_{22})^2$$

$$= [X_{11}^2, \sqrt{2} X_{11} X_{12}, X_{12}^2] \begin{bmatrix} X_{21}^2 \\ \sqrt{2} X_{21} X_{22} \\ X_{22}^2 \end{bmatrix}$$

$$= \phi(X_i) \times \phi(X_j)$$

eg: $\left([1, \sqrt{2}, 1] \begin{bmatrix} 4 \\ \sqrt{2}\,4 \\ 4 \end{bmatrix} \right)$ $= 4 + 8 + 4 = 16$

5

MANY KERNELS

POLYNOMIAL KERNEL,
GAUSSIAN KERNEL

$K(x_i, x_j) = \mu - x_i$

AMAZING WHAT'S THE CATCH

1 SHATTERING A SET: — SLIDES

ALL POSSIBLE WAYS OF

"AN SET OF INST SHATTERS S is

shattered by hypothesis space H
iff ~~for~~ every dichotomy of S there
exists some

2 HOW TO CHOOSE THE RIGHT KERNEL

KERNEL ALIGNMENT

$\underset{L}{\arg\min}\left(K(x_i, x_j), Y \cdot Y^T\right)$

ALL THE SAME LABEL
~~ONE~~ SIM 1, OTHER
0