TIML - Assign 1 Solns.

15-Feb-19

(1) We need 'B' such that: $P[w^T X - b \geq 0] \geq B$

i.e. 'B' such that $P[b - w^T X > 0] \leq 1 - B$

$$P[\underbrace{b - w^T X > 0}_{(*)}] = P\left[-\frac{2}{3}X_1 + X_2 > 0\right]$$

→ Our bounds answer the upper bound query; hence this trick

$$= P\left[-\frac{2}{3}(X_1 - 2) + \left(X_2 - \frac{3}{4}\right) > \frac{7}{12}\right]$$

say, Z

→ Our bounds work on mean zero r.v.'s hence this trick.

$$= P\left[e^{\lambda Z} > e^{\lambda 7/12}\right] \quad \forall\, \lambda > 0$$

$$\leq \frac{E[e^{\lambda Z}]}{e^{7\lambda/12}} \longrightarrow \text{Markov Inequality}$$

$$= \frac{E\left[e^{-\frac{2\lambda}{3}(X_1 - 2)}\right] E\left[e^{\lambda(X_2 - 3/4)}\right]}{e^{7\lambda/12}} \quad (\because X_1 \text{ is ind. of } X_2)$$

Now,

$0.5 \leq X_1 \leq 2.5 \Rightarrow -\frac{1}{3} \leq -\frac{2}{3}(X_1 - 2) \leq 1 \Rightarrow E\left[e^{-\frac{2}{3}\lambda(X_1 - 2)}\right] \leq e^{\lambda^2 2/9}$

$0.5 \leq X_2 \leq 1.5 \Rightarrow -\frac{1}{4} \leq X_2 - 3/4 \leq 3/4 \Rightarrow E\left[e^{\lambda(X_2 - 3/4)}\right] \leq e^{\lambda^2/8}$

Hoeffding bound

$$\therefore \quad (*) \leq e^{2\lambda^2/9}\, e^{\lambda^2/8}\, e^{-7\lambda/12} \quad \forall\, \lambda > 0$$

Chernoff bounding technique

$$\Rightarrow (*) \leq e^{\min_{\lambda} 2\lambda^2/9 + \lambda^2/8 - 7\lambda/12}$$

$$\Rightarrow (*) \leq e^{-0.245}$$

$$= e^{-0.245} \approx 0.7827$$

$$\Rightarrow P[w^T X - b > 0] \geq 1 - e^{-0.245} \approx 0.2173$$

* Given only the mean of the test datapoint one would label it to be a positive datapoint. But the prob. that it is indeed such one is can

(2) Using hints given in lectures, Smola book we arrive at following claim:

claim $$P\left[\sup_{f\in\mathcal{F}}\left\{R_{emp}[f]-R[f]\right\}>\varepsilon\right]\longrightarrow 0 \text{ as } m\to\infty \quad \forall\varepsilon>0$$
(Required nec. & suff. cond.)

$$\Downarrow$$

$$\sup_{f\in\mathcal{F}} R_{emp}[f] \xrightarrow{P} \sup_{f\in\mathcal{F}} R[f] \qquad \text{(Emp. unif. max.)}$$

i.e. $$P\left[\left|\sup_{f\in\mathcal{F}} R_{emp}[f]-\sup_{f\in\mathcal{F}} R[f]\right|>\varepsilon\right]\longrightarrow 0 \text{ as } m\to\infty$$

Proof:

$$P\left[\left|\sup_{f\in\mathcal{F}} R_{emp}[f]-\sup_{f\in\mathcal{F}} R[f]\right|>\varepsilon\right] = P\left[\sup_{f\in\mathcal{F}} R_{emp}[f]-\sup_{f\in\mathcal{F}} R[f]>\varepsilon\right] + P\left[\sup_{f\in\mathcal{F}} R[f]-\sup_{f\in\mathcal{F}} R_{emp}[f]>\varepsilon\right]$$

$$\geq \left(\text{Let } f^m=\operatorname{argmax}_{f\in\mathcal{F}} R_{emp}[f]\ ;\ f^{ort}=\operatorname{argmax}_{f\in\mathcal{F}} R[f]\right)$$

$$= P\left[R_{emp}[f^m]-R[f^{ort}]>\varepsilon\right] + P\left[R[f^{ort}]-R_{emp}[f^m]>\varepsilon\right]$$

$$= P\left[R_{emp}[f^m]-R[f^m]+\underbrace{R[f^m]-R[f^{ort}]}_{\leq 0}>\varepsilon\right] + P\left[R[f^{ort}]-R_{emp}[f^{ort}]+\underbrace{R_{emp}[f^{ort}]}_{\leq 0}-R_{emp}[f^m]>\varepsilon\right]$$

$$\leq P\left[R_{emp}[f^m]-R[f^m]>\varepsilon\right] + P\left[R[f^{ort}]-R_{emp}[f^{ort}]>\varepsilon\right]$$

$$\leq P\left[\sup_{f\in\mathcal{F}}\left\{R_{emp}[f]-R[f]\right\}>\varepsilon\right] + P\left[R[f^{ort}]-R_{emp}[f^{ort}]>\varepsilon\right]$$

$\downarrow m\to\infty$ by one-sided unif. conv. cond. given

$\downarrow m\to\infty$ by Chernoff's bound.

0 $\qquad$ 0

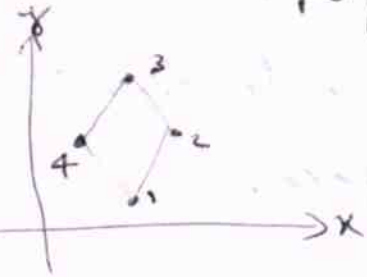
* In other words, the one-sided uniform conv. are necessary for ERMin. & ERMax. & hence, the two-sided (regular) uniform conv. conditions are necessary for both ERMin & ERMax. *

(3) Refer pg. 40 (Thm. 1 & Proof) in appendix of Bulgers Tutorial.

(4) Let $\mathcal{F}_{11} \equiv$ set of all axis ||'l ellipses
$\mathcal{F}_{45} =$ set of all ellipses at 45°
$\mathcal{F} = \mathcal{F}_{45} \cup \mathcal{F}_{11}$.

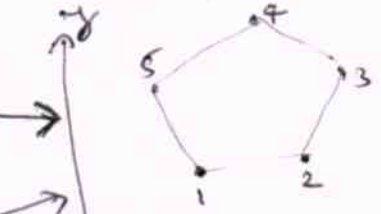
(Everywhere we are drawing 'regular' polygons)

It is easy to see that $\mathcal{F}_{11}$ shatters



square ~~regular~~

also $\mathcal{F}$ shatters →



regular pentagon.

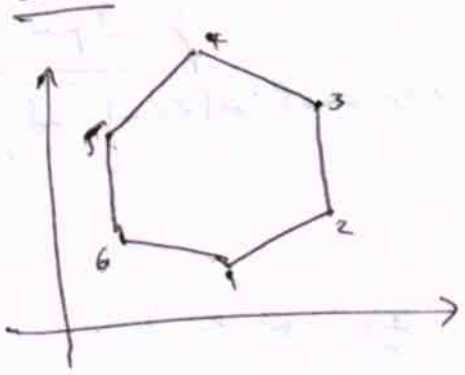
$\mathcal{F}_{11}$ does not shatter →

(In none save regular polygons in optimal arrangements; we need to look at orientation with axis through for this problem)

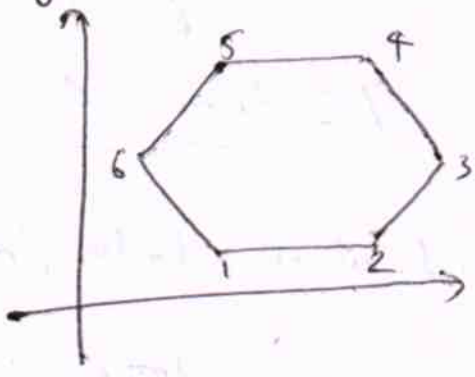
Hence $\mathcal{F}_{11}$ is VC dim. = 4 (see also rem. to next problem)

$\mathcal{F}$ does not shatter

regular hexagons in both orientations



and



Hence $\mathcal{F}$ is VC dim. = 5

* It is easy to verify all claims regarding shattering made here by imagining ellipses to be limiting cases of ~~ellipses~~ rectangles *

(5) Refer to paper titled "Classification by Polynomial Surfaces" by Martin Anthony

Ans

Simply use the fact illustrated in lecture that polynomial discriminants are essentially linear discriminants in expanded feature space.

For $\mathcal{F}_{11}$:

$$f(x) = \text{sign}\left((x-c)^T S (x-c) - 1\right) = \text{sign}\left(\begin{bmatrix} x_1 - c_1 & x_2 - c_2 \end{bmatrix} \begin{bmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{bmatrix} \begin{bmatrix} x_1 - c_1 \\ x_1 - c_2 \end{bmatrix} - 1 \right)$$

$$= \text{sign}\left(\omega_1 z_1 + \omega_2 z_2 + \omega_3 z_3 + \omega_4 z_4 + \omega_5 \right)$$

where $\omega_1 = \delta_1$, $\omega_2 = \delta_2$, $\omega_3 = -2\delta_1 c_1$, $\omega_4 = -2\delta_2 c_2$, $\omega_5 = \delta_1 c_1^2 + \delta_2 c_2^2 - 1$ ↓ fixed if $\omega_1, \omega_2, \omega_3, \omega_4$ are fixed.

& $z_1 = x_1^2$, $z_2 = x_2^2$, $z_3 = x_1$, $z_4 = x_2$

This is essentially linear discriminants in $\mathbb{R}^4$ passing through fixed point

$\therefore$ VC dim = 4 (not 5 since constant is fixed)

(6) Bound which is to be plotted is:

$$\begin{cases} 4 \exp\{ m \log 2 - m\varepsilon^2/8 \} & \text{for } m \leq 3 \\ 4 \exp\{ 8\left(\log \frac{m}{3} + 1\right) - m\varepsilon^2/8 \} & \text{for } m > 3 \end{cases}$$

Assignment 3 Solutions

Section II

Q2 In 3-d, the VC dim. of hyperplanes is 4, which is upper bound on that for canonical hyperplanes

The margin margin achievable with 4 points lying on a sphere is when they are on a tetrahedron (regular).

Height of tetrahedron $= \frac{4}{3}R$ $\left(\because \text{margin} = \frac{2}{\|w\|} \leq \frac{4R}{3}\right)$

$\therefore$ If $\|w\| \geq \frac{3}{2R}$ then V.C dim $= 4$

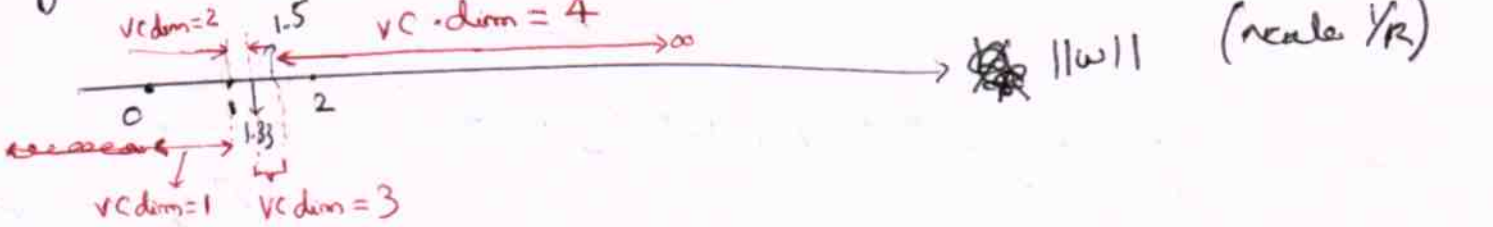
Next possibility is still higher margin where 3 pts. can be shattered
$\rightarrow$ pts. lying on an equilateral Δle with circumradius $= R$.
$\rightarrow$ Height of this eq. Δle $= \frac{3R}{2}$ $\left(\because \text{margin} = \frac{2}{\|w\|} \leq \frac{3R}{2}\right)$

$\therefore$ If $\frac{4}{3R} \leq \|w\| < \frac{3}{2R}$, then VC dim $= 3$

III by next possibility is 2 pts. & margin $\leq 2R$

$\Rightarrow$ If $\frac{1}{R} \leq \|w\| < \frac{4}{3R}$, then VC dim $= 2$.

ofcourse if $\|w\| < \frac{1}{R}$, then V dim $= 1$.



Q3 SVM primal with sq hi. loss:

$$\min_{w,b,\xi} \quad \frac{1}{2}\|w\|^2 + C\sum \xi_i^2$$

$$\text{s.t.} \quad y_i(w^T x_i - b) \geq 1 - \xi_i, \quad \xi_i \geq 0.$$

$$\mathcal{L} = \frac{1}{2}\|w\|^2 + C\sum \xi_i^2 + \sum \alpha_i [1 - y_i w^T x_i + y_i(b) + 1 - \xi_i] - \sum \beta_i \xi_i$$

$\alpha_i, \beta_i \geq 0$ are Lagrange mul.

$\nabla_w \mathcal{L} = 0 \Rightarrow w = \sum \alpha_i y_i x_i$

$\frac{\partial \mathcal{L}}{\partial \xi_i} = 0 \Rightarrow \alpha_i + \beta_i = 2C\xi_i$ ← (only change compared to hinge-loss)

$\frac{\partial \mathcal{L}}{\partial b} = 0 \Rightarrow \sum \alpha_i y_i = 0.$

$$\mathcal{L} = -\frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum \alpha_i - \sum_i \frac{(\alpha_i + \beta_i)^2}{4C}$$

(eliminating w, b, ξ_i)

∴ Dual is $\max_{\alpha_i, \beta_i} -\frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i^T x_j - \sum_i \frac{(\alpha_i + \beta_i)^2}{4C} + \sum \alpha_i$

s.t. $\sum \alpha_i y_i = 0, \quad \alpha_i \geq 0, \quad \beta_i \geq 0$

Same as $\max_{\alpha_i} -\frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum \alpha_i + \underbrace{\max_{\beta_i \geq 0} -\sum_i \frac{(\alpha_i + \beta_i)^2}{4C}}$

s.t. $\sum \alpha_i y_i = 0, \alpha_i \geq 0$

$\Rightarrow \beta_i = 0 \ \forall i$

Same as $\max_{\alpha_i} -\frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i^T x_j - \sum_i \frac{\alpha_i^2}{4C} + \sum \alpha_i$

s.t. $\alpha_i \geq 0, \ \sum_i \alpha_i y_i = 0$

Can be merged ↓

Same as $\max_{\alpha_i} -\frac{1}{2}\sum \alpha_i \alpha_j y_i y_j k_2(x_i, x_j) + \sum \alpha_i$

s.t. $\alpha_i \geq 0, \ \sum_i \alpha_i y_i = 0$

where $k_2(x_i, x_j) = x_i^T x_j + \frac{1}{2C}$ if $i = j$; $x_i^T x_j$ if $i \neq j$

Realize that exactly same as hard-margin SVM dual

Only difference is gram matrix is incremented by $\frac{1}{2C} I$

Dual is also equal to: $-\min_{\alpha} \frac{1}{2}\alpha^T Q_c \alpha - 1^T\alpha$ (D)

(vectorial notation) s.t. $\alpha \geq 0$, $y^T\alpha = 0$ → vector of ones of dim l's

where y is vector of labels of tr. datapts.

Y is $\text{diag}(y)$.

$Q_c = Y\left(K + \frac{1}{2c}I\right)Y$

K → gram matrix of tr. datapts

Note that version of (D) is Q_c which is strictly positive definite ∀ c

∴ the objective is strictly convex & hence solution is unique.

→ recovery of w & b:

w can be (uniquely) computed using $w = \sum \alpha_i y_i x_i$.

Once we get w, look at those eg. where $\alpha_i > 0$ let them be set S → set of support vectors.

$\forall\, i \in S$, $\xi_i = 1 + y_i b - y_i w^T x_i \geq 0$

ξ_i: complementary slackness cond.; $\xi_i \geq 0$

say set S_+

$\Rightarrow$ $\forall\, i \in S$, $y_i = 1$, we have $b \geq w^T x_i - 1$

$\forall\, i \in S$, $y_i = -1$, we have $b \leq w^T x_i + 1$

say set S_-

$$\max_{i \in S_+} w^T x_i - 1 \leq b \leq \min_{i \in S_-} w^T x_i + 1$$

Q4 There are many ways of showing this is an indefinite kernel.

Here is one simple way:

If k is positive then Cauchy-Schwartz inequality must hold

i.e. if Cauchy-Schwartz ineq. does not hold then k is not positive.

In our case $k(x,y)^2 \le k(x,x)k(y,y)$

would imply $(x^T y)^2 \le 0 \quad \forall x, y$ which is impossible

(~~contradicts the fact that the zero vector cannot~~)

$\therefore$ k is not positive. Hence $-k$ is also not positive. $\therefore$ It is an indefinite kernel.

Q5 We will show that $k(A_1, A_2) = \frac{|A_1 \cap A_2|}{|\Omega|}$ is a positive kernel. then of course $3e^{|A_1 \cap A_2|^2}$ is also positive kernel.

[This kernel is actually very close to what we have seen in lecture]

Consider the prob. space $\mathbb{P} = (\Omega, 2^\Omega, P)$ → P is "uniform distribution" or "classical probability"

i.e. $P(A) = \frac{|A|}{|\Omega|}$.

In this $\mathbb{P}$, we have $P(A_1 \cap A_2) = \frac{|A_1 \cap A_2|}{|\Omega|} = k(A_1, A_2)$.

Consider the indicator r.v.: ~~mapping~~ ~~$\Omega \to \mathbb{R}$~~

$X_{A_1} = \mathbb{1}_{A_1}$ (then $E[X_{A_1}] = P(A_1) = \frac{|A_1|}{|\Omega|}$)

Here, $E[X_{A_1} X_{A_2}] = P(A_1 \cap A_2) = k(A_1, A_2)$

↓

we already know that, $E[XY]$ is a valid inner product in

Note that $E\{X^2\} \geq 0 \quad \forall X$ (Non-negativity)

& $E\{X^2\} = 0 \Rightarrow X = 0$ with prob 1

& $E\{XY\} = E\{YX\}$ (symmetry)

& $E\{(\alpha_1 X_1 + \alpha_2 X_2) Y\} = \alpha_1 E\{X_1 Y\} + \alpha_2 E\{X_2 Y\}$ (linearity)

Hence $E\{XY\}$ is indeed a valid inner product.

Here $k(A_1, A_2) = \frac{|A_1 \cap A_2|}{|\Omega|}$ is a positive kernel.

In fact the vector space of all rv's defined on $\mathbb{P}$ is an inner-product space where k is an inner-product.

Another argument: Consider the kernel $k(A_1, A_2) = |A_1 \cap A_2|$

Now we will see "another" feature space where this is an inner-product.

Consider a Euclidean space in $|\Omega|$ dimensions.
(one axis for each element of Ω)

For each $A \in 2^{\Omega}$, consider the mapping of A to a binary vector of 0, 1's such that a '1' is put for all elements in A.

In this space, of course the natural dot product between vectors formed from A_1 & A_2 will give $|A_1 \cap A_2|$

(6) Refer Scholkopf, Smola, Williamson, Bartlett. New Support Vector Algorithms, Neural Computation, 12, pp. 1207-1245, 2000.

(7) Since it is mentioned that it is a real-world regression problem we can safely assume that the loss is bounded.

$$a \le (y - f(w))^2 \le b \quad \text{—— } (1/2 \text{ mark})$$

Now the learning bounds we derived in lectures using R.A. do hold for all bounded loss functions. Here the problem boils down to simply upper bounding the conditional R.A. of the induced set of loss functions:

$$\mathcal{R}_m(\mathcal{L}) = E_\sigma \left[\sup_{\substack{\|w\| \le W \\ \|x_i\| \le R}} \left\{ \frac{1}{m} \sum_{i=1}^{m} \sigma_i (y_i - w^T x_i)^2 \right\} \right] \quad \text{—— } (1/2 \text{ mark})$$

Now bounding $\mathcal{R}_m(\mathcal{L})$ is not easy. ($1/2$ mark) will be given to any attempt (even if wrong) towards it. Here one (incomplete) way of bounding based on derivations in lectures:

$$\mathcal{R}_m(\mathcal{L}) \le E_\sigma \left[\sup_{\substack{\|w\| \le W \\ \|x_i\| \le R}} \left\{ \frac{1}{m} \sum_i \sigma_i y_i^2 \right\} \right] + E_\sigma \left[\sup_{\substack{\|w\| \le W \\ \|x_i\| \le R}} \left\{ \frac{2}{m} \sum_i \sigma_i w^T x_i \right\} \right]$$

$$+ E_\sigma \left[\sup_{\substack{\|w\| \le W \\ \|x_i\| \le R}} \left\{ \frac{1}{m} \sum_i \sigma_i (w^T x_i)^2 \right\} \right]$$

$$\le \frac{2W}{m} \sqrt{\text{trace}(XX^T)} \quad \text{(using ideas given in lecture)}$$

$$\le \frac{W^2}{m} E_\sigma \left[\left(\text{max. eigenvalue of } \sum_{i=1}^{m} \sigma_i x_i x_i^T \right)^2 \right]$$

This bound is fine if $\to 0$ as $m \to \infty$. Can we ensure that it happens?

Coming to the crux of a completely correct answer to the problem, refer to Cor. 6.7 in "Local Rademacher Complexities" by Bartlett, Bousquet & Mendelson, Annals of Statistics, 33(4): 1497-1537, 2005.

(8) Refer to related wikipedia articles.

Applications: (a) generalizations of CLT
(b) Efron-Stein inequality (Jackknife)
(c) Used in lectures for proving McDiarmid ineq.

(9) $f(X) = \text{sign}\left(\overbrace{1.5(X_1-1)^2 + 1.5(X_2-1)^2 + (X_1-1)(X_2-1) - 1}^{-g(X_1,X_2)} \right)$.

We require $P[-g(X_1,X_2) > 0] \geq ?$ → there is a typo in the problem!! we need a lower bound instead of upper bound

We have $E\{g\} = +\left(1 - 1.5 - 1.5 \times 7.5 - (0.5)(-1.5)\right) = -5$

We can get $P[g > 0] \leq ?$ using McDiarmid's inequality of Hoeffding's inequality.

McDiarmid $P\{g>0\} = P[g - E\{g\} > 5] \leq e^{-2(5)^2 / \sum_{i=1}^{2} c_i^2}$ → b.d. property

Now $\underset{X_1, X_1'}{\text{sup}} |g(X_1,X_2) - g(X_1',X_2)| = \underset{X_1, X_1'}{\text{sup}} \left| (X_1' - X_1)\left[1.5(X_1' + X_1) + X_2 - 4\right] \right|$

$\leq \begin{cases} 44 \rightarrow \text{obvious bound} \\ 21 \rightarrow \text{through mat lab.} \end{cases} = c_1$

Ully $c_2 = 12$

$\therefore P[f = 1] \geq 1 - e^{-2(5)^2/(c_1^2 + 12^2)} \simeq 0.08192$

Hoeffding's Ineq.

Again $P\{g>0\} = P\{\overbrace{g - E[g]}^{Z} > s\}$

$= P\{e^{\lambda Z} > e^{\lambda s}\} \leq \frac{E\{e^{\lambda Z}\}}{e^{\lambda s}}$

We can show that

$-19.5 \leq Z \leq 6$

$\leq e^{\frac{\lambda^2}{8}(25.5)^2 - s\lambda}$ $\forall \lambda > 0$

$\therefore P\{\hat{H}=1\} \geq 1 - e^{-s^2/(25.5)^2} = 0.07401$

Again, since McDiarmid inequality employs information that the R.Vs X_1, X_2 are independent, it wins!