SMAI-M20-L28: Illustrative Numerical Problems (Cont.)

C. V. Jawahar

IIIT Hyderabad

October 21, 2020

Class Review

Backpropagation

1. Convergence
2. Relationship between Loss and Acccuracy
3. Accuracy on Train Set and Test Set
4. Characterizing Optimization Problem in MLP

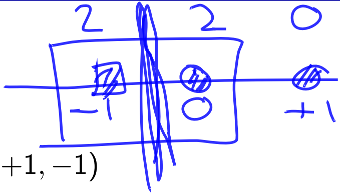
Blank

Recap:

- **Supervised Learning:** Formulation, Conceptual Issues, Concerns etc.
- **Classifiers:** (i) Nearest Neighbour, (ii) Notion of a Linear Classifier (iii) Perceptrons (iv) Bayesian Optimal Classifier (v) Logistic Regression (vi) Multiclass classification architectures (v) SVMs (hard margin, soft margin, kernel)(vi) MLP (vii) BP
- **Dimensionality Reduction and Applications:** (i) Feature Selection and Extraction (ii) PCA (iii) LDA (iv) Eigen face
- **Matrix Factorization and Applications:** (i) SVD, (ii) Eigen Decomposition (iii) Matrix Completion (iv) LSI (v) Recommendations
- **Other Topics:**
  - Linear Regression
  - Probabilistic View, Bayesian View, MLE
  - Gradient Descent: Stochastic and Batch GD
  - Loss Functions and Optimization
  - Eigen Vector based optimization
  - Neuron model, Single Layer Perceptrons
  - Kernel Functions and Kernel Matrix

Blank

Illustrative Problem - I (Done: Q?)



Consider a linearly separable data

$$(-1,+1),(0,-1),(+1,-1)$$

$x = -1/2$

1. Write the primal problem. Geometrically show the feasible region in a w, b 2D plane. Show the optimal w as -2 and b as -1. Geometrically validate the solution.

   $wx + b = 1$

2. Write the dual objective $J(\alpha_1, \alpha_2, \alpha_3)$. And the constraint $\sum_i \alpha_i y_i = 0$. Show that the solution is $\alpha_1 = \alpha_2 = 2$ and $\alpha_3 = 0$. Which are the support vectors then?
3. Show that the dual solutiion also leads to the same w and b.

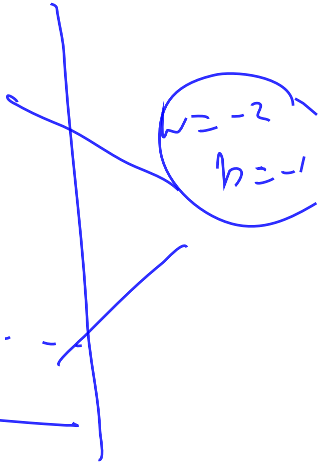
$w = \sum_{1} \alpha_1 y_1 x_1$ $\quad$ $w = -2 \quad b = -1$

- Primal

  $\min \|w\|$

- Dual

  $\max \sum_{i} \alpha_i - \sum\sum \alpha_i \cdots$



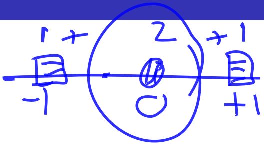
Blank

Blank

Blank

Illustrative Problem -II (Done; Q?)

Consider a linearly in-separable 1D data



$$(-1,+1),(0,-1),(+1,+1)$$

Demonstrate how the problem can be solved with a quadratic kernel $(\mathrm{p}^T q+1)^2$

if $x^2 < 0.5$ -1
else $+1$

1. Write the dual objective with the constraint $\sum_i \alpha_i y_i = 0$
2. Substitute for α_2 from the constraint in the objective. (i.e., eliminate α_2)
3. Differentiate wrt α_1 and α_3 and equate to zero. Solve to obtain $\alpha_1 = \alpha_3 = 1$ and $\alpha_2 = 2$ => All three are SV
4. For a new sample the decision is

$2x^2 - 1 = 0$
$x^2 = 1/2$

$$sign(\sum_i \alpha_i y_i \kappa(x_i, x) + b)$$

$x_i^T x$

Simplify and get this expression as $2x^2 - 1$. (Assume $b = -1$. Indeed, you can find b yourself.)

$K(x_i, x)$

Blank

$$\text{sign}(w^T x + b) - \text{lin}$$

$$\boxed{\text{sign}\left(\sum_{i=1}^{\cancel{N}\ |sv|} \alpha_i y_i K(x_i, x) + b\right)}$$

$$w = \sum_{i=1}^{N} \alpha_i y_i x_i$$

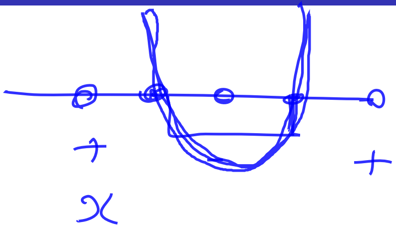
$|sv| = 100$

100 times

slow

Blank

$$2x^2 - 1 = 0$$



$$x^2 = 1/2$$

$$\underline{\underline{x \longrightarrow f(x)}}$$

0	0	0
-1	0	$+1$

Blank

Blank

Illustrative Problem - III

x_1, x_2

Consider the ExOR Problem:

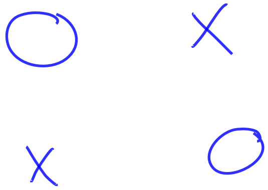
x_1	x_2	y
-1	-1	-1
-1	+1	+1
+1	-1	+1
+1	+1	-1

$+1 \quad (p^T p + 1)^2 = 9$

$-1 \quad (p^T q + 1)^2 = 1$

-1

$+1$



1. Consider a kernel $(p^T q + 1)^2$ and construct the Kernel matrix
2. Write the dual objective in terms of α_1, α_2, α_3 and α_4.
3. Differentiate wrt α_1, α_2, α_3 and α_4 and equate to zero. You get four equations.
4. Solve (guess) the values for α_1, α_2, α_3 and α_4. Same?
5. We know $\phi() = [1, x_1^2, \sqrt{2}x_1x_2, x_2^2, \sqrt{2}x_1, \sqrt{2}x_2]^T$. Find w as $\sum_i \alpha_i y_i \phi(x_i)]$. Does it solve the exor? why?

Blank

$$\begin{bmatrix} 9 & 1 & 1 & 1 \\ 1 & 9 & 1 & 1 \\ 1 & 1 & 9 & 9 \\ 1 & 1 & 1 & 9 \end{bmatrix} 4 \times 4 \qquad \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \underbrace{x_i^T x_j}_{K(x_i, x)}$$

$$\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 - \frac{1}{2}\Big(9\alpha_1^2 - 2\alpha_1\alpha_2 - 2\alpha_1\alpha_3$$

$$+ 2\alpha_1\alpha_4 + 9\alpha_2^2 + 2\alpha_2\alpha_3 - 2\alpha_2\alpha_4$$

$$+ 9\alpha_3^2 - 2\alpha_3\alpha_4 + 9\alpha_4^2 \Big)$$

$$9\alpha_1 - \alpha_2 - \alpha_3 + \alpha_4 = 1$$

$$-\alpha_1 + 9\alpha_2 + \alpha_3 - \alpha_4 = 1$$

$$-\alpha_1 + \alpha_2 + 9\alpha_3 - \alpha_4 = 1$$

$$\alpha_1 - \alpha_2 - \alpha_3 + 9\alpha_4 = 1$$

$$\boxed{\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \frac{1}{8}}$$

$$\phi = [1,\ x_1^2,\ \sqrt{2}x_1x_2,\ x_2^2,\ \sqrt{2}x_1,\ \sqrt{2}x_2]^T$$

$$\begin{matrix} -1 & -1 \\ -1 & +1 \end{matrix}$$

$$w = \sum \alpha_i y_i \phi(x_i)$$

$$w = \frac{1}{8}\left[-1\begin{bmatrix} 1 \\ 1 \\ \sqrt{2} \\ 1 \\ -\sqrt{2} \\ -\sqrt{2} \end{bmatrix} + 1\begin{bmatrix} 1 \\ 1 \\ -\sqrt{2} \\ 1 \\ -\sqrt{2} \\ \sqrt{2} \end{bmatrix} \cdots \right] = \begin{bmatrix} 0 \\ 0 \\ -1/\sqrt{2} \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$\underline{\sqrt{2}x_1x_2}$

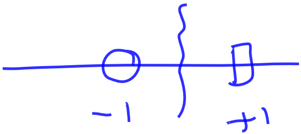
$\underline{w^T\phi(x)}$

Blank

x_1	x_2	y
-1	-1	-1
-1	$+1$	$+1$
$+1$	-1	$+1$
$+1$	$+1$	-1

$\phi() \Rightarrow$

$x_3 = x_1 x_2$	y
$+1$	-1
-1	$+1$
-1	$+1$
$+1$	-1



Blank

$$\mathrm{sign}\left(\sum_{i=1}^{|SV|} \alpha_i y_i K(x_i, x) - b\right)$$

Blank

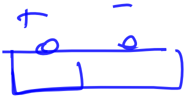
Blank

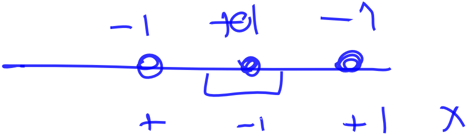
Illustrative Problem - IV

VC dimension of a function class is the largest set that can be shattered by a member of the function class.

1. Interval $[\alpha_1, \alpha_2]$ on real line. (i.e., +ve is inside $[\alpha_1, \alpha_2]$ else negative) 2
2. Axis parallel rectangles in 2D. (i.e., positive inside and negative outside) 4
3. Convex Polygons in 2D (inside positive and outside negative)
4. Union of intervals in 1D.

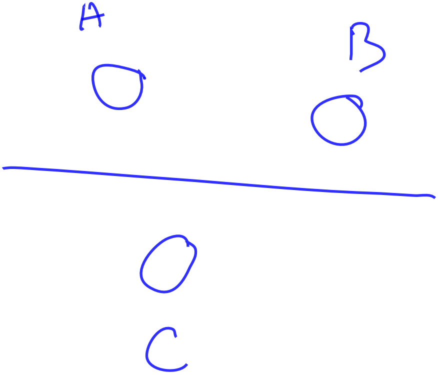
[1]





[1] VC Dimension is d, if a set of size d can be shattered completely and no set of size $d + 1$ can be shattered.

Blank



Set = 3 member

F. Class = {lin}

Blank

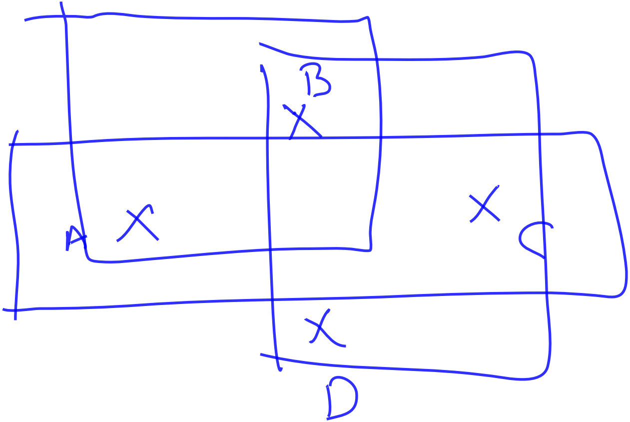
$$\boxed{\begin{cases} + & \text{if } x > \alpha \\ - & \text{else} \end{cases}} \to 1$$

+ −

o •

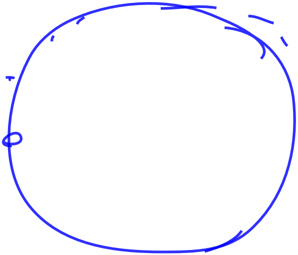
− +

✗ ✓



Any set of 5 can not be drawn

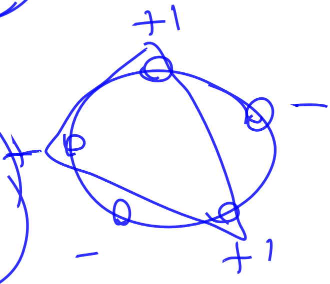
Blank



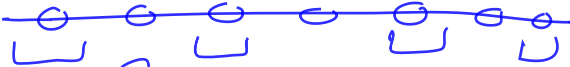
100

VC Dim d

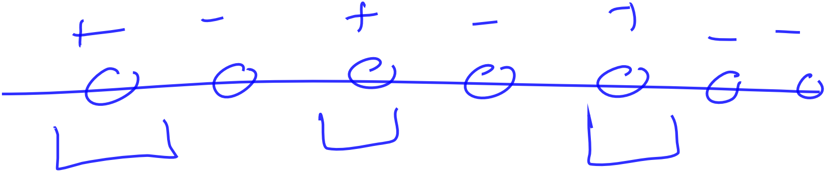
Convex pol $= \infty$



Blank



$$\bigcup_i [\alpha_i, \alpha_0]$$



Blank

Blank

Discussion Point (Repeat, if time permits)

(Advanced; Out of Syllabus!!) We know that a new kernel can be defined in terms of existing kernels:

$$\sum_{i=1}^{K} \alpha_i \kappa(\cdot, \cdot)$$

Then why don't we formulate the overall learning problem in SVM, including that of learning these α_i

1. Discuss why it is a good idea?
2. How do we use it for "fusing" different features?
3. Why do we limit to $\sum$?

See some of the works relevant[2] and [3]. Read later.

[2] http://manikvarma.org/pubs/varma07c.pdf

[3] https://cvit.iiit.ac.in/images/ConferencePapers/2009/Rakesh09More.pdf

Blank

Blank

Discussion Point (Repeat, if time permits)

(Advanced; Out of Syllabus!!)
We know that linear SVMs are superefficient (compared to K-SVMs).
Can we find a $\phi()$ corresponding to a Kernel and solve the problem as

$$\mathbf{w}^T \phi(x)$$

Indeed, this may become difficult for many kernels (eg. RBFs). **why?**
Can we find a finite dimensional approximation of $\phi()$? How does it help in speeding up SVM with no major reduction in accuracy?
read [4] and [5] later.

1. Discuss why it is a good idea?
2. Suggest an application where speed matters (eg. in the reference is that ofobject detection).

[4] https://cvit.iiit.ac.in/images/ConferencePapers/2010/Sreekanth10Generalized.pdf
[5] https://www.robots.ox.ac.uk/ vgg/publications/2011/Vedaldi11/vedaldi11.pdf

Blank

Blank

What Next:?

Friday Lecture Session:

1. No formal online session
2. Submit the solution to the five problems we solved in class (11am on Friday). Write neatly with sufficient explanations and equations/pictures/sketches/details.

Next?

1. NN Architectures and NN Learning
2. Programming for Deep Learning.
3. Beyond Supervised Learning