SMAI-M20-L27: Illustrative Numerical Problems

C. V. Jawahar

IIIT Hyderabad

October 19, 2020

Class Review

MLP

1. Number of parameters
2. What can it solve?
3. What can it guarantee?

VC Dimension

Blank

Recap:

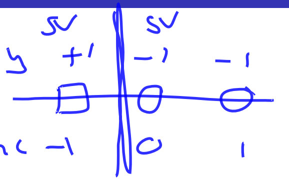
- **Supervised Learning:** Formulation, Conceptual Issues, Concerns etc.
- **Classifiers:** (i) Nearest Neighbour, (ii) Notion of a Linear Classifier (iii) Perceptrons (iv) Bayesian Optimal Classifier (v) Logistic Regression (vi) Multiclass classification architectures (v) SVMs (hard margin, soft margin, kernel)(vi) MLP
- **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

Consider a linearly separable data

x y x_2 y_2 x_3 y_3

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



$wx+b=0$

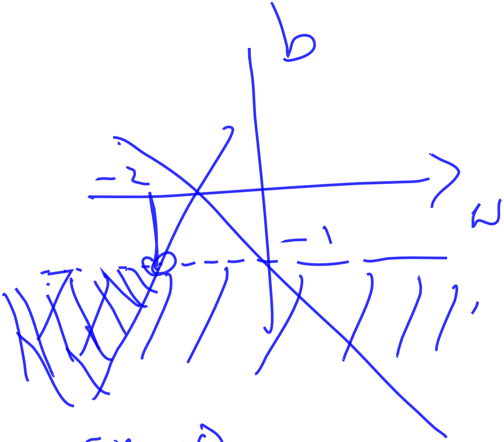
1. Write the primal problem. Geometrically show the feasible region in a w, b 2D plane. Show the optimal w as -2 and b as -1. Geometrically validate the solution.
2. Write the dual objective $J(\alpha_1, \alpha_2, \alpha_3)$. And the constraint $\sum_i \alpha_i y_i = 0$. Show that the solution is $\alpha_1 = \alpha_2 = 2$ and $\alpha_3 = 0$. Which are the support vectors then?
3. Show that the dual solutiion also leads to the same w and b.

Blank

$$\min \frac{1}{2}\|w\|$$

$$y_i(w^T x_i + b) \geq 1$$

$$\min \frac{1}{2} w^2$$



$$+1(w(-1)+b) \geq 1$$

$$\boxed{-w+b \geq 1} \quad 1$$

$$(w^* \; b^*) = (-2, -1)$$

$$-2x + -1 = 0$$

$$\boxed{x = -1/2}$$

Blank

$$\sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{d} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$$\alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2}\left(\alpha_1^2 + \alpha_3^2 - 2\alpha_1\alpha_3(-1)\right)$$

$$\boxed{\alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2}(\alpha_1 + \alpha_3)^2} \quad — 1$$

$$\sum_i \alpha_i y_i = 0 \implies \alpha_1 - \alpha_2 - \alpha_3 = 0$$

$$\alpha_2 = \alpha_1 - \alpha_3 \quad — 1$$

$$\boxed{2\alpha_1 - \frac{1}{2}(\alpha_1 + \alpha_3)^2} \implies \alpha_3 = 0 \quad —$$

$$\boxed{wx + b = +1}$$

3

$$2\alpha_1 - \frac{1}{2}\alpha_1^2$$

$$\boxed{\max -\frac{1}{2}(\alpha_1 - 2)^2 + 2}$$

$\alpha_1 = 2$

$$\alpha_1 = 2 \quad \alpha_2 = 2 \quad \alpha_3 = 0$$

$$\boxed{w = \sum_i \alpha_i y_i x_i}$$

$w = -2$

$b = -1$

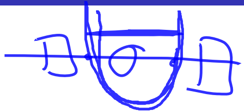
Blank

Blank

Blank

Illustrative Problem -II

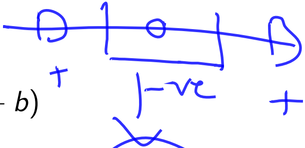
Consider a linearly in-separable 1D data



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

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

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 $\alpha_2 = 2$
4. For a new sample the decision is



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

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

Blank

$$\sum_{i=1}^{N} \alpha_i - \sum_i \sum_j \alpha_i \alpha_j y_i y_j \boxed{K(x_i, x_j)}$$

$$K = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 4 \end{bmatrix}$$

$$K(p, q) = (\bar{p} q + 1)^2$$

-1, 0, $+1$

Blank

$$\alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2}\left(4\alpha_1^2 + \alpha_2^2 + 4\alpha_3^2 - \alpha_2\alpha_1 - \alpha_2\alpha_3 \ldots\right)$$

$$\alpha_1 - \alpha_2 + \alpha_3 = 0 \qquad \sum_i \alpha_i y_i = 0$$

$$\alpha_2 = \alpha_1 + \alpha_3$$

$$2\alpha_1 + 2\alpha_3 - \left(2\alpha_1^2 + 2\alpha_3^2 - \frac{1}{2}(\alpha_1 + \alpha_3)^2\right)$$

Blank

$\frac{\partial}{\partial \alpha_1} = 0$

$$2 - 3\alpha_1 + \alpha_3 = 0 \qquad \frac{\partial}{\partial \alpha_3} = 0$$

$$2 + \alpha_1 - 3\alpha_3 = 0$$

$(*)$

$$\boxed{\alpha_1 = 1 \qquad \alpha_3 = 1 \qquad \alpha_2 = 2}$$

Blank

$$y_i(w^T x + b) \geq +1$$

$$\boxed{y_i(w^T + b) = +1}$$

Illustrative Problem - III

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. Axis parallel rectangles in 2D. (i.e., positive inside and negative outside)

[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

Blank

Illustrative Problem - IV

Consider the ExOR Problem:

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

Blank

Blank

Blank

Blank

Blank

What Next:? (next three)

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