SMAI-M20-L15: Perceptron Algorithm

C. V. Jawahar

IIIT Hyderabad

September 11, 2020

Class Review

$$\Sigma = \frac{1}{N}\sum_{i=1}^{N}[x_i-\mu][x_i-\mu]^T$$

$$\Sigma = \frac{1}{N}\sum_{i=1}^{N} x_i x_i^T$$

Propeties of SVD

- What are the properties of U, D, V?
- How SVD and EV are related?
- What do we say about SVD of square matrix?
- What can we say about rank deficiency and SVD?

$$\begin{array}{cc} & N \\ d & \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \end{array}$$

$$\underset{d \times d}{XX^T} \qquad \underset{N \times N}{X^TX}$$

Announcements

1. **Serious Observation**
   - **Unethical collaborations and Practices**:
     - **Open** collaboration. If you want to create any unofficial channels/models/groups make sure that a TA or official representative is there. Talk to the instructor. Take permission.
     - Reposting/sharing class videos/slides/notes without permission
   - Please inform these/similar and anything else that is going on and take corrective actions in next 48 Hrs.

Blank

Recap:

- Problem Space:
    - Learn a function $y = f(\mathbf{W}, \mathbf{x})$ from the data.
    - Dimesnionality Reduction and Representation (Feature Selection, PCA, Neural Embeddings)
    - Matrix Factorization for Data Matrices: (LSI, Matrix Completion, Recommendation Systems)
- Supervised Learning:
    - Notions of Training, Validation and Testing; Loss Function and Optimization
    - Generalization, Overfitting, Occam's razor, Model Complexity, Bias and Variance, Regularization.
    - Performance Metrics, Estimating error using validation set.
- Algorithms:
    - Nearest Neighbour, Linear Classification; Linear Regression
    - Optimal Decision as ω_1 if $P(\omega_1|\mathbf{x}) \geq P(\omega_2|\mathbf{x})$ else ω_2
    - PCA, Eigen Face
    - Gradient Descent Optimization

This Lecture:

$(1, 1)$ $+1$
$(-1\ -1)$ -1

1. MSE using GD (Delta Rule)
   - $\mathbf{w}^{k+1} \leftarrow \mathbf{w}^k + \eta y_i \mathbf{x}_i$
   - Appreciate that perceptrons and this are different.
2. Variations in GD (Stochastic and Mini Batch)
   - Single sample, mini batch, batch
   - Stochastic versions
3. Neuron Models
   - $y = \phi(\mathbf{w}^T \mathbf{x})$
   - Step, sigmoid, tanh etc.

Questions? Comments?

Blank

Discussion Point - I

$w^{k+1} \leftarrow w^k - \eta \nabla J$

We know there are better update rules than gradient descent?

1. Write the newton's update rule?
2. Why is still Newton's method not preferred? [1]

0.2 ?

$d \times d$

$n \times n$

[1] https://stats.stackexchange.com/questions/253632/why-is-newtons-method-not-widely-used-in-machine-learning

Blank

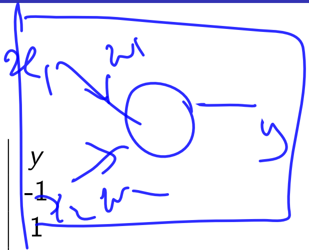
Blank

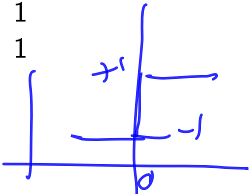
Discussions Point - II

We know the AND and OR logic in $\{-1, +1\}$ as

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

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



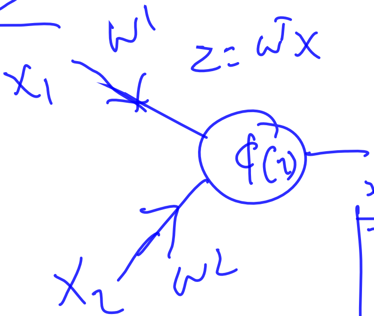


1. Which is AND? which is OR?
2. Design a two input neuron with $\phi(z)$ as *sign*(z) for both AND and OR. Draw pictorially.
3. Can we do NAND and NOR similarly? (Try later)
4. Can we do ExOR? (draw and see). Is it Linearly seperable? Ans: NO

$$\mathrm{sig}(z) = \begin{cases} +1 & \text{if } z \geq 0 \\ -1 & \text{else } z < 0 \end{cases}$$

AND

-1	-1	-1
-1	1	-1
-1	-1	-1
1	1	1



$w1 = 1$

$w2 = 1$

x_1	x_2	y	z	$\phi(z)$
-1	-1	-1	-2	-1

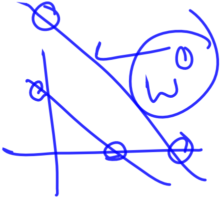
Blank

Discussions Point - III

Consider the following three samples and their labels $((x_1, x_2), y)$:

S1 S2 S3

$$\{((1,1),+), \quad ((2,2),-), \quad ((0,0),+)\}$$



Look at the perceptron update rule with $\eta = 0.1$

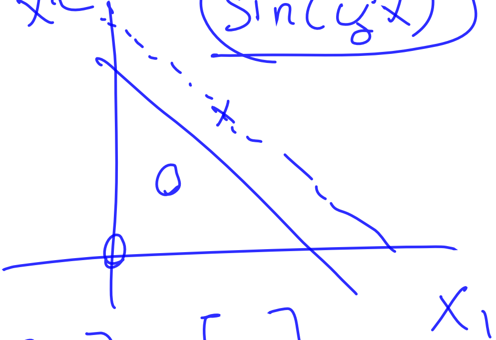
$$\mathbf{w}^{k+1} \leftarrow \mathbf{w}^k + \eta \sum_{\mathbf{x}_i \in \mathcal{E}} y_i \mathbf{x}_i$$

Start with line equations given below and do two iterations. Did it converge? If not, how many more iterations will it take?

$x_1 + x_2 = 4$

- line $x_1 = x_2$
- line that pass through (0,2) and (2,0)
- line that pass through (0,4) and (4,0)

$\sin(w^T x)$



$x_1 + x_2 = 4$

$x_1 + x_2 - 4 = 0$

$$\underbrace{\begin{bmatrix} 1 & 1 & -4 \end{bmatrix}}_{w^T} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix}}_{x}$$

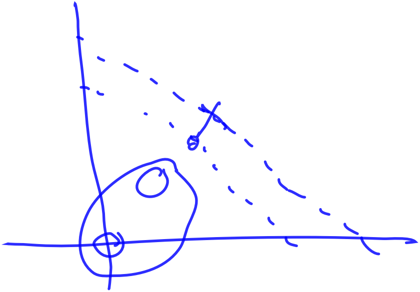
$$\underbrace{\begin{bmatrix} 1 \\ 1 \\ -4 \end{bmatrix}}_{w^0} \rightarrow \underbrace{\begin{bmatrix} \\ \\ \\ \end{bmatrix}}_{w^1}$$

$\cancel{w_0^T x}$

Blank

$$w' \leftarrow \begin{bmatrix}1\\1\\-4\end{bmatrix} + 0.1\left((+1)\begin{bmatrix}0\\0\\1\end{bmatrix} + (+1)\begin{bmatrix}1\\1\\1\end{bmatrix}\right)$$

$$\leftarrow \begin{bmatrix}1\\1\\-3.8\end{bmatrix}$$



Blank

Blank

What Next:?

1. More about Gradient Descent
2. Neuron Model and Perceptrons
3. Analysis of Perceptron Algorithm