SMAI-M20-L21: Introduction to Kernels

C. V. Jawahar

IIIT Hyderabad

September 28, 2020

Announcements

- **Quiz 1**
    - The same time as last week.
    - The same set of topics we planned.
- **Course Evaluation**
    - We may have to minorly tweak the course evaluation models.
    - Will be announced next week.

Class Review

We are given a problem of Multi-Class Classification with

- DDAG
- BHC
- One-vs-Rest.

Now we want to know:

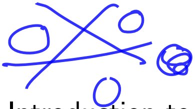
1. What is the computational advantages of each architecture?
2. How many nodes (binary classifiers) are required?
3. When is the decision ambiguous?
4. How do we compute the accuracy of the system from accuracy of the individual classifiers?

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
- **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
    - MLE
    - Gradient Descent
    - Stochastic and Batch GD
    - Eigen Vector based optimization
    - Neuron model
    - Loss Functions and Optimization

This Lecture:

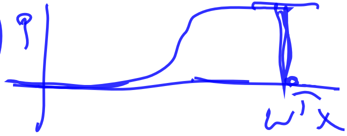


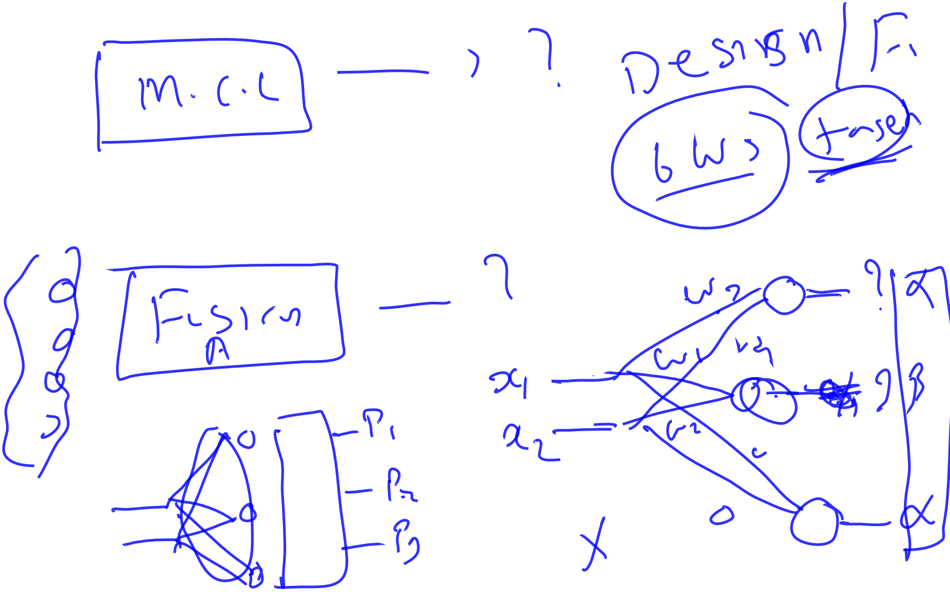
$$\frac{e^{w_i^T x}}{\sum e^{w_i^T x}} \qquad \max_i w_i^T x$$

1. Introduction to Kernels
   - Kernel Trick: A method of solving nonlinear problems with linear algorithms.
   - Kernel Function: $\kappa(p, q) = \phi(p)^T \phi(q)$
2. Extending the idea of Logistic Regression to Multi-Class
   - Classifiers output a score that a decision. Making fusion simpler.
   - Softmax: Find the maximum, normalize and probabilistic interpretation.

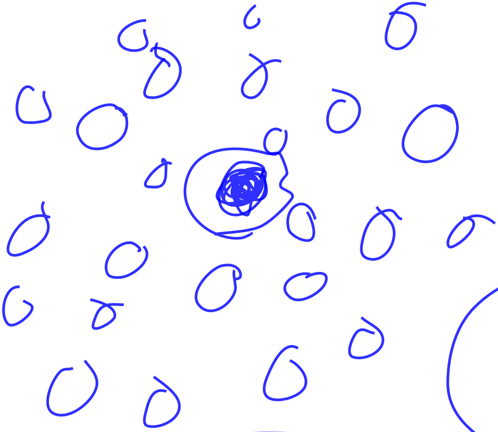
Questions? Comments?

$\boxed{\mathrm{Sign}(c^T x)}$ $\qquad w^T x$




M.C.L
?
Design
GWs
Fusion
A
?
x_1
x_2
w_2
w_1
P_1
P_2
P_3
?
α
β

Blank



100

any pair ~100%

pairwise better than multiclass

$100C_2$ 100

$w^T x$

$\phi(w)^T \phi(x)$

$K(w, x)$

We had seen the Kernel

$$\kappa(\mathbf{p}, \mathbf{q}) = \phi(\mathbf{p})^T \phi(\mathbf{q})$$

with

$$\kappa(\mathbf{p}, \mathbf{q}) = (\mathbf{p}^T \mathbf{q})^2$$

This means:

$$\boxed{\phi : \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} \rightarrow \begin{bmatrix} p_1^2 \\ p_2^2 \\ \sqrt{2} p_1 p_2 \end{bmatrix}}$$

homoge-
nous kern

$(p^T q + 1)^2$

Is the feature map unique given the kernel?

Blank

$$\phi_1 \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} \longrightarrow \begin{bmatrix} p_1^2 \\ p_2^2 \\ \sqrt{2} p_1 p_2 \end{bmatrix} \qquad 2D \longrightarrow 3D$$

$$\phi_2 \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} \Longrightarrow \begin{bmatrix} p_1^2 \\ p_1^2 \\ \sqrt{2} p_1 p_2 \end{bmatrix} \qquad 2D \longrightarrow 3D$$

$$\begin{bmatrix} p_1 \\ p_2 \end{bmatrix} \longrightarrow \begin{bmatrix} p_1^2 \\ p_2^2 \\ p_1 p_2 \\ p_2 p_1 \end{bmatrix} \qquad 2D \longrightarrow 4D$$

- $K(\quad)$ Kernel eg. $(p^T q)^2$
- ϕ : Feature Map

No need to know ϕ.
Existence of ϕ is good enough

Blank

Discussions Point -II

quadratic kernels

We now the $\phi()$ correspond to:

$$\kappa(\mathbf{p}, \mathbf{q}) = (\mathbf{p}^T \mathbf{q})^2$$

(homogene.)

2D -> 3D

What is the $\phi()$ correspond to:

$$\kappa(\mathbf{p}, \mathbf{q}) = (\mathbf{p}^T \mathbf{q} + 1)^2$$

(nonho...)

2D -> 6D

(assume $\mathbf{p} \in R^2$)

2

Blank

$$\left(\begin{bmatrix} p_1 \\ p_2 \end{bmatrix}^T \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} + 1\right)^2 = \left(p_1 q_1 + p_2 q_2 + 1\right)^2 \qquad (a+b+c)^2$$

$$\left[p_1^2 q_1^2 + p_2^2 q_2^2 + 1^2 + \qquad \right]$$

$$\begin{bmatrix} p_1^2 \\ p_2^2 \\ \vdots \\ \vdots \end{bmatrix}^T \begin{bmatrix} q_1^2 \\ q_2^2 \\ \vdots \\ \vdots \end{bmatrix}$$

$$\left[p_1^2 \quad p_2^2 \quad p_1 p_2 \quad p_1 \quad p_2 \quad 1 \right]$$

Blank

$$e^{-\beta J q} = \underline{\text{inf series}}$$

$$\tanh(\beta J q)$$

Blank

Discussion Point - III

$x_1, \ldots, x_N$ "Kernel Matrix"

Assume there are N samples.

A kernel matrix $\mathbf{K}$ is defined as a matrix with (i, j)th element as the kernel computed with the i th and j th sample. i.e.,

$$K_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j)$$

1. What is the dimension of the Kernel Matrix?
2. Is Kernel matrix square?
3. Is Kernel matrix symmetric?
4. Is Kernel matrix PSD?

"Y"

$N \times N$

$\kappa(p, q) = (p^T q)$

Blank

For what class of funch

$k(\quad)$ "valid kernel"

there exist a ϕ?

$k(\quad) = (pq)^{\sim} \quad (q+1)^{\sim}$

Eq.

What Next:? (next three)

1. Winding up (i) Logistic Regression (ii) Multi-Class Classification and (iii) LDA
2. SVMs and Kernels