SMAI-M20-L25: Nonlinear methods: SVM, Kernels and MLP

C. V. Jawahar

IIIT Hyderabad

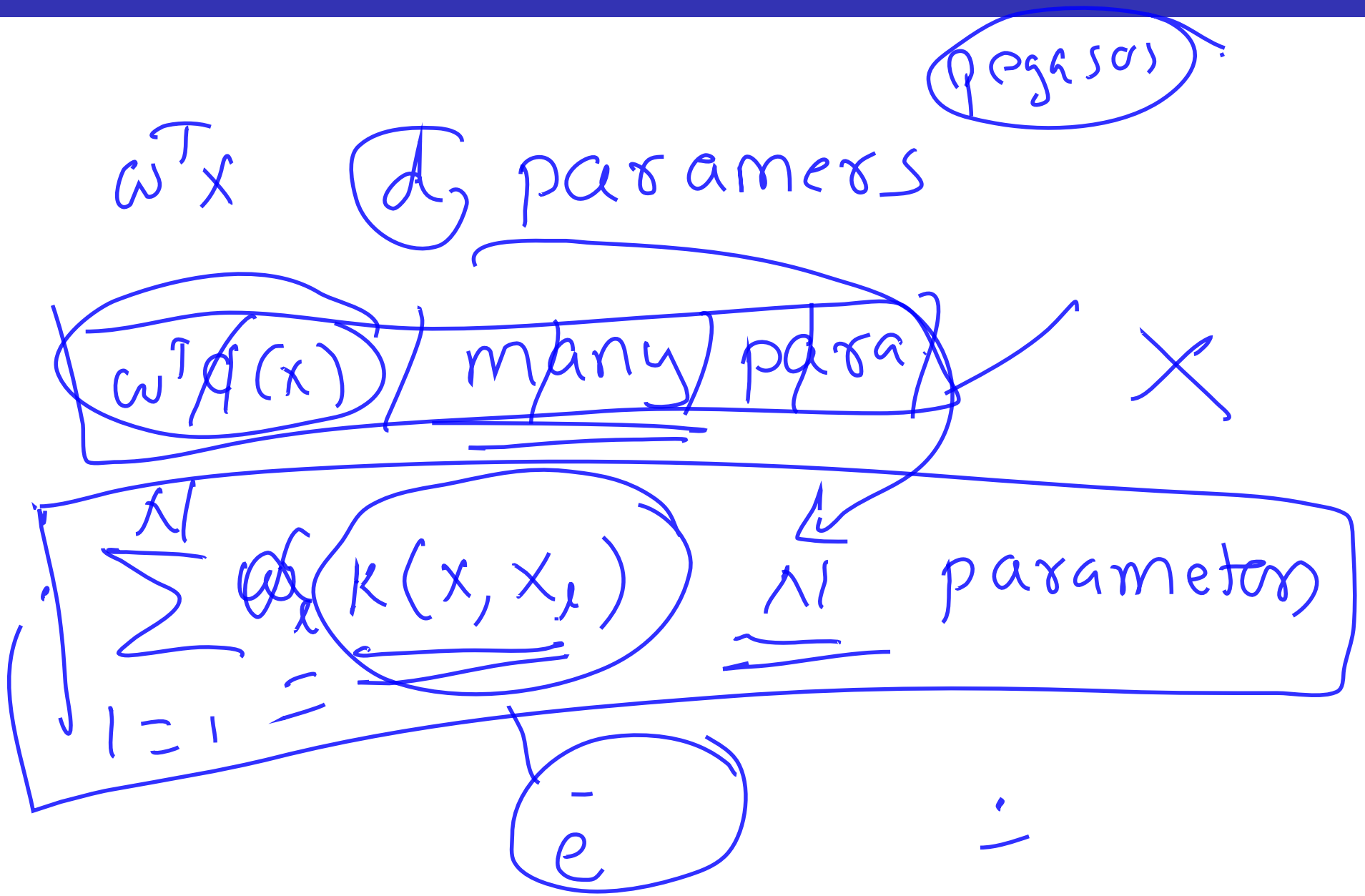
October 12, 2020

Class Review

Questions is based on the brief review of Kernels you had seen at: `https://www.dropbox.com/s/qryziuo3u143q5e/KERNEL-REVIEW.pdf?dl=0`(shared in the class last week).
(see specially Section 4 for today)

1. Relationship between Data Matrix and Kernel Matrix.
2. How computations such as (i) centering (ii) distance (iii) normalization can be done in the feature space ($\phi()$).

Blank


pegasos
$w^T x$
d parameters
$w^T \phi(x)$
many para
X
$\sum_{i=1}^{N} \alpha_i k(x, x_i)$
N
parameters
e

Recap:

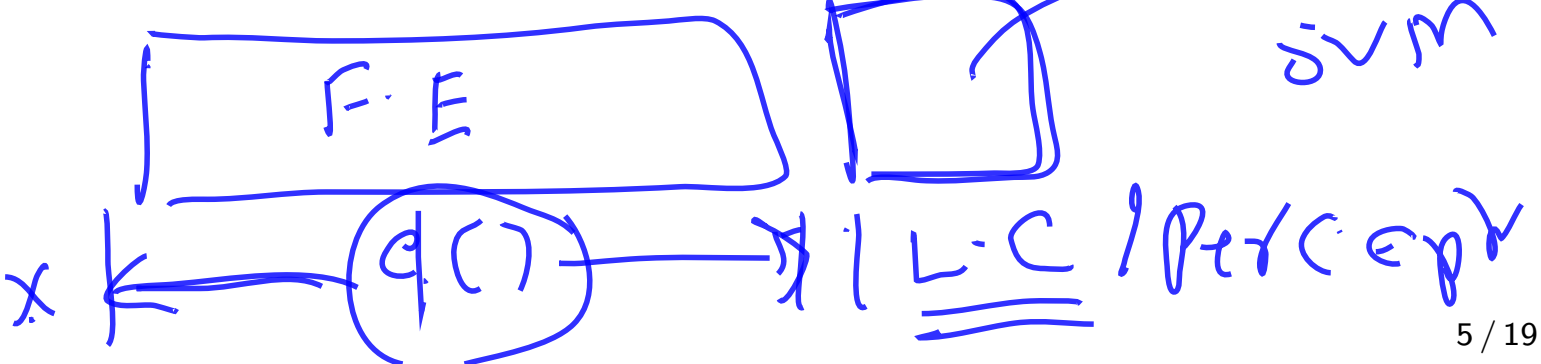
- **Supervised Learning:** Formulation, Conceptual Issues, Concerns etc.
- **Classifiers:** (i) Nearest Neighbour, (ii) Notion of a Linear Classifier (iii) Perceptrons (iv) Bayesian Optimal Classifier (v) Logistic Regression (vi) Multiclass classification architectures (v) SVMs (hard margin, soft margin,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

$K(x, y) = (x^T y)^r$

This Lecture:

1. Kernel SVMs:
   - What happens during training?
   - What happens during testing?
2. MLP:
   - Introduction to a typical MLP architectures
   - Appreciating "Deep MLP" (aka Deep Neural Network) as feature transformation followed by a classification.

Questions? Comments?



Blank

Blank

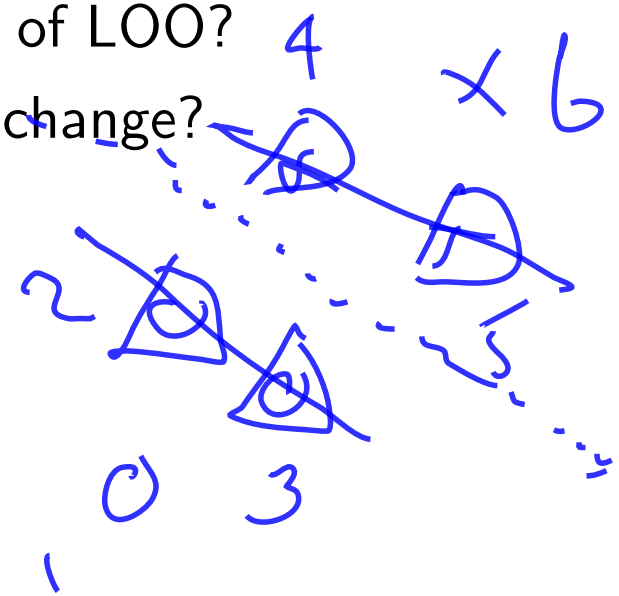
Blank

Discussions Point -I

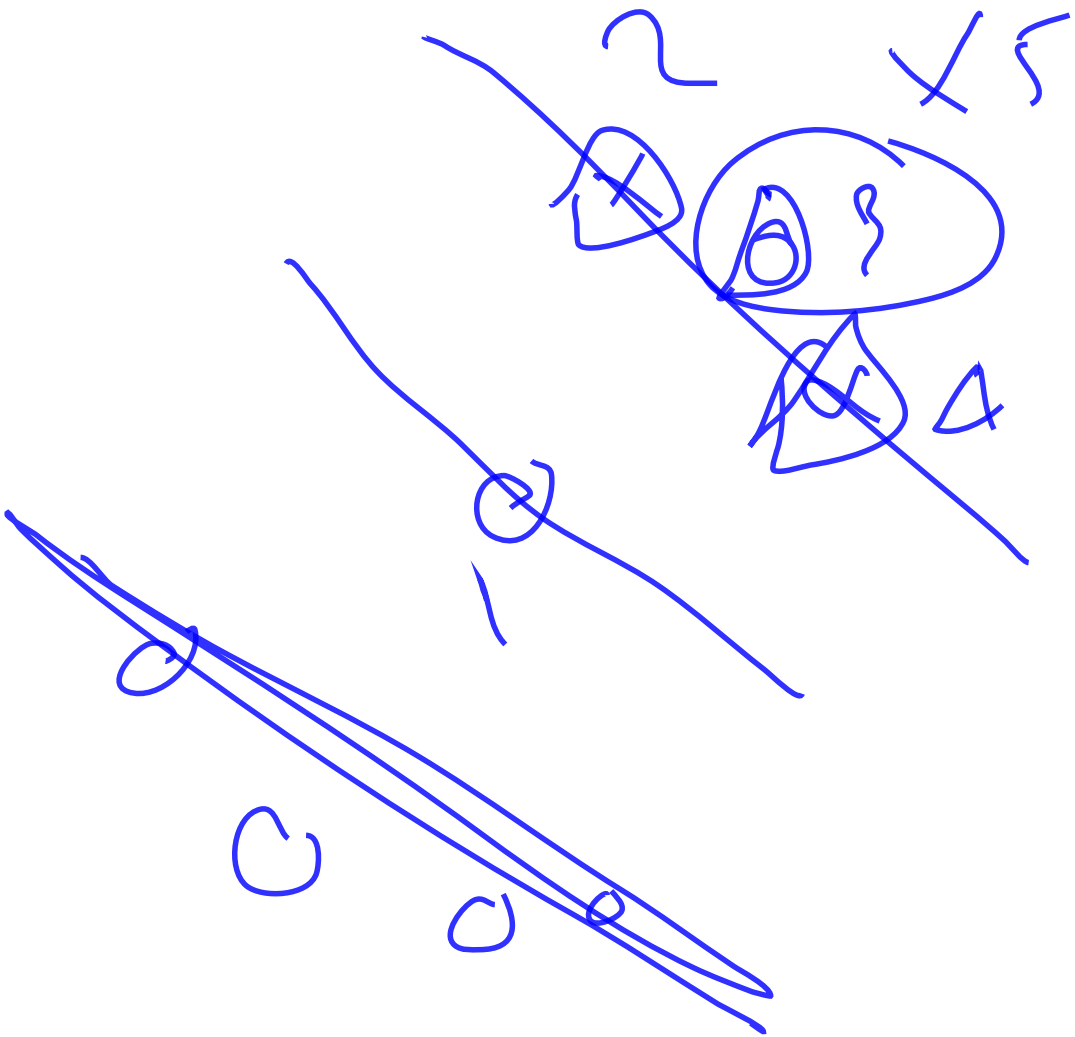
"Leave One Out"

A typical SVM training leaves you with the set of Support Vectors (SV) and the corresponding αs (for all others αs are zero).

1. Comment: "The ratio $\frac{|SV|}{N}$ gives us an idea of the error rate." Why? Argue?
2. Is this an upper bound or lower bound of LOO?
3. If we change the kernel, does the SVs change?



Blank



Blank

$x_1 \cdots x_N$

$\{x_1 \cdots\} = SV$

$\{\alpha_i\}$

$w^T x$ $O(d)$ 1

100

$\sum_{i=1} \alpha_i y_i K(x_i, x)$

$(x_i^T x)^2$

$O(d)$

Blank

Discussions Point -II

Starting from the Lagrangian

$$L(\mathbf{w}, b, \alpha, \xi) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \frac{C}{2}\sum_{i=1}^{N}\xi_i^2 - \sum_{i=1}^{N}\alpha_i[y_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 + \xi_i]$$

i.e., Derive the dual function for L2 SVM
(Write on a paper, complete and submit later.)

$\frac{\partial L}{\partial \xi_1} = 0$

$\frac{C}{2} 2\xi_1 - \alpha_1$

$C\xi_1 - \alpha_1 = 0$

Blank

$$C \xi_i - \alpha_i = 0 \quad \forall i$$

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

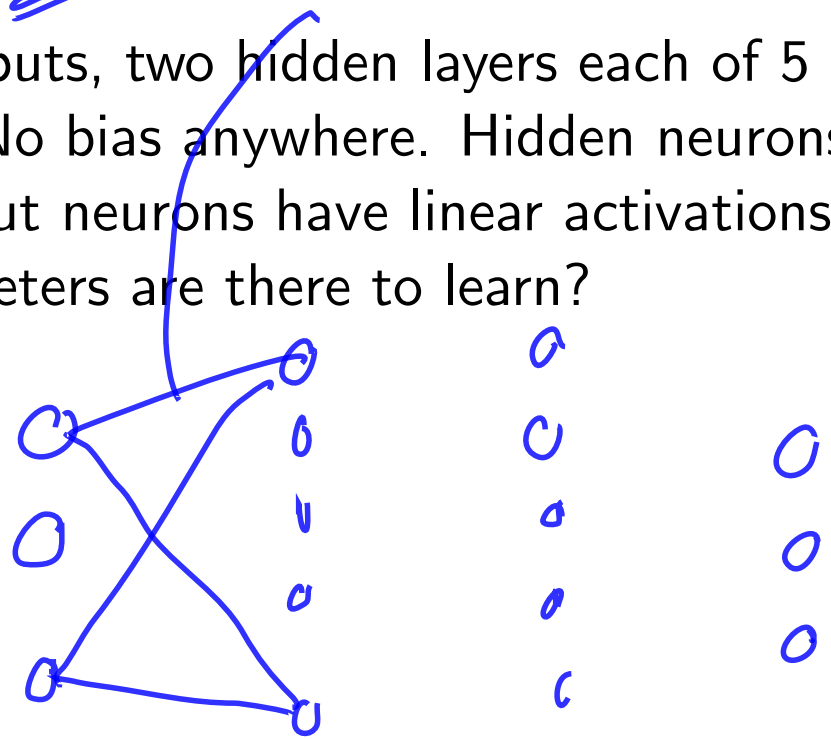
Blank

Blank

Blank

Discussion Point - III

An MLP has 3 inputs, two hidden layers each of 5 neurons, and three output neurons. No bias anywhere. Hidden neurons have sigmoid activations. Output neurons have linear activations.
How many parameters are there to learn?



What Next:? (next few)

1. SVMs and Kernels (winding up)
2. MLP and Backpropagation