SMAI-M20-L26:Nonlinear methods: SVM, Kernels and MLP

C. V. Jawahar

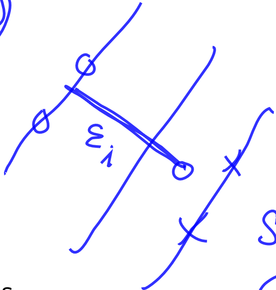
IIIT Hyderabad

October 16, 2020

"tight"

$\boxed{1 \geq \alpha}$

KKT

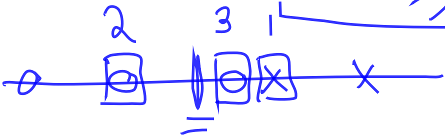


- Hard Margin SVM
- Soft Margin SVM
- Kernel SVM
- Specific Questions based on lecture notes.

Properties and limiting cases.

small

$\boxed{\geq 1-\varepsilon_i}$



Blank

$$\boxed{\text{Can we prune s.v?}}$$

$$w = \sum_{i=1}^{|Sv|} \alpha_i x_i = \sum_{i=1}^{|Sv|/2} \beta_i x_i$$

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 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

This Lecture:

1. MLP Architecture
    1. Role of Activations
    2. Regression, Classification and choice of output neurons.
    3. Expressive power of neural networks.
2. Chain rule for computing gradients
    1. How gradients can be computed
    2. What should we keep in mind while defining the layers.
3. Backpropagation through chain rule.
    1. Appreciate how BP works
    2. Why "back" in the BP ?
4. Kernel Ridge Regression
    1. Another example of Kernelization
    2. Familiarity of K and Φ

Questions? Comments?

Blank

Discussions Point - I

(0, 1, ..., 100)

100

0 w_1
1 w_2
2 w_3

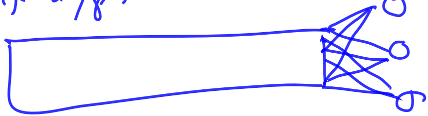
We have a three-class classification problem. We want to use an MLP

A we can use a single output neuron and force it to output 0, 1 and 2 corresponding to the three classes.

B we can have three output neurons with classes coded as [1,0,0], [0,1,0] and [0,0,1].

1. Which one will you prefer? Why?

② what shall be the activation of o/p?



w_1	w_2	w_3
1	0	0
0	1	0
0	0	1

Blank

Blank

Blank

Discussions Point -II

$$(BA + \lambda I)(BA + \lambda I)^{-1} B = (BA + \lambda I) B (A + \cdot)^{-1}$$

Kernel Ridge Regression:

1. We used the result

$$(BA + \lambda I)^{-1} B \ = B(AB + \lambda I)^{-1}$$

   verify this.

2. What are the steps in the training time for K-Ridge regression? What. are the steps during testing?

Blank

Train

Input:

(x_i, y_i) $i = 1 \cdots N$, Kernel K

o/p

α_i $i = 1 - N$

Testing

Test

$\underline{\underline{x_i}}$, $\underline{\underline{\alpha_i}}$, $\underline{\underline{\boxed{x}}}$

o/p

$$\boxed{y} = \sum_{i=1}^{N} \alpha_i K(x_i, x)$$

Blank

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

learn

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

Mult K~ Le

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 $\sum$?

$$K = \prod_{i=1}^{K} \alpha_i$$

See some of the works relevant[1] and [2]. Read later.

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

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

Blank

Blank

Discussion Point - IV

(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

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

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? ϕ'
read [3] and [4] later.

$$\boxed{w^T \phi'(x)}$$

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

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

Blank

Blank

What Next:? (next three)

1. SVMs and Kernels (winding up?)
2. NN Architectures and NN Learning
3. Programming for Deep Learning.
4. Next Lecture (Mon or Wed): Problem Solving related to SVM and Kernels. (no new video content!).