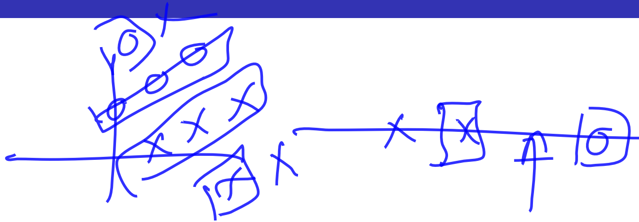
# SMAI-M20-L24: SVM and Kernels

C. V. Jawahar

IIIT Hyderabad

October 9, 2020



$x_1 \quad \phi(x_1) \quad \phi(x_n)$

- SVM, Support Vectors and when do they change and how does it affect the margin.
- Kernels in SVM. How are the computations done?

$w^T x$

all samples $\sum_{i=1}^{N} \alpha_i$

$\sum_{i=1}^{N} \alpha_i y_i x_i$

# Blank

Orig Alg

$\omega^T x$

$O(d)$

Re-write

$w = \sum_{i=1}^{N} \beta_i x_i$

$d \to N$

$\omega \quad \beta$

$\sum_{i=1}^{N} \beta_i x_i^T x$

$O(N)$

Q

$x \to \phi(x)$

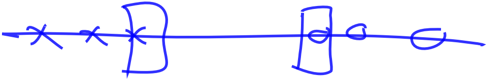
$\sum_{i=1}^{N} \beta_i \phi(x_i)^T \phi(x)$

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

# 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
- **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:



$\square \ \alpha \neq 0$

1. Kernels:
   - Examples of PD Kernels (or valid kernels)
   - Kernels on structures (or non-vector) data
2. SVM:
   - Derivation of Dual from the Primal

**Questions? Comments?**

$$\boxed{K(\underline{x_i}, \underline{x_j})}$$

# Blank

# Discussions Point -I

We know the perceptron classification as $\bar{w}^T x$ – orig

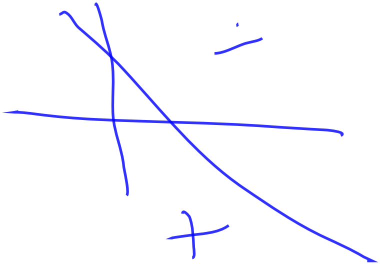
$$\boxed{sign(\sum_{i=1}^{N} \alpha_i \mathbf{x}_i^T \mathbf{x}))}$$ — rewritten

and the kernel perceptron as

$$sign(\sum_{i=1}^{N} \alpha_i \kappa(\mathbf{x}_i, \mathbf{x}))$$ — Ker. form

1. Does the kernel perceptron yield a nonlinear boundary? Yes
2. Assume the samples were in 2D, how do we plot (or visualize the decision boundary)? ?

# Blank





$$w^T x = 0$$

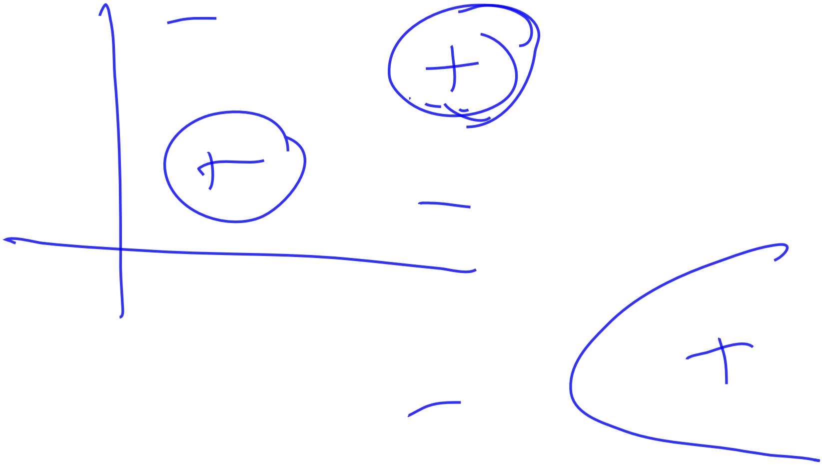
Percep

$w^T x$ $\quad$ $\sum \alpha_i x_i^T x$

K-Percep

$\sum \alpha_i K(x_i, x)$

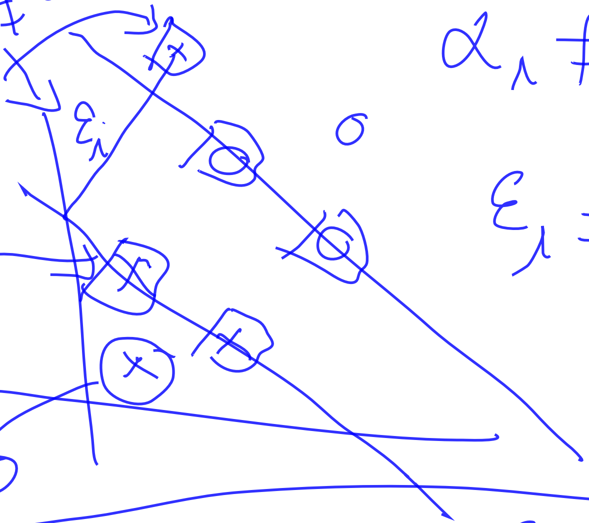
# Blank



$\alpha_i \neq 0 \Rightarrow SV$

$\xi_i = 0$

$\alpha_i \neq 0$, $\xi_i \neq 0$



$\alpha_i \neq 0$

$\xi_i = 0$

$\alpha_i = 0$

$\xi_i = 0$

$$\boxed{\alpha_i \neq 0 \ \&\& \ \xi_i = 0}$$

# Blank

# Discussions Point -II

$\text{sign}(w^T x + b)$

We had seen how that

$$\boxed{\mathbf{w}^* = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i}$$

$b$?

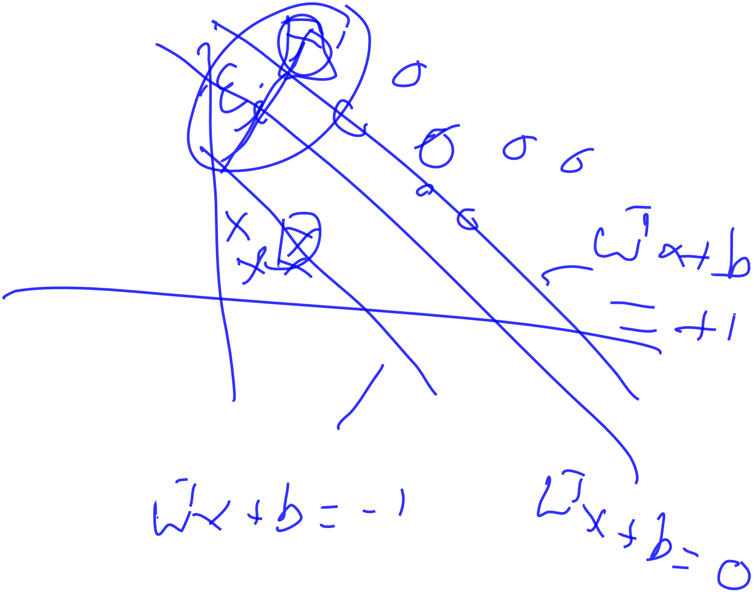
1. How do we find $b^*$ for hard margin SVM?
2. How do we find $b^*$ for soft margin SVM? [1]
3. Support Vectors (SV) are the vectors where $\alpha_i$s non-zero. Why $\alpha$ zero for non-support vector? [2]

$$J_d(\alpha) = \sum_{i=1}^{N} \alpha_i - \sum \sum \alpha_i \alpha_j$$

[1] https://stats.stackexchange.com/questions/451868/calculating-the-value-of-b-in-an-svm

[2] https://stats.stackexchange.com/questions/355661/svm-why-alpha-for-non-support-vector-is-zero

# Blank



# Blank

$D = 500 +,\ 500 -$

opt $\underline{\underline{J_d(\alpha)}} \Rightarrow \alpha^*$ $N$

$\boxed{\alpha_i \neq 0 \Rightarrow SV}$

$\underline{\underline{10\ SV}}$

# Discussion Point - III

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.)

# What Next:? (next few)

1. SVMs and Kernerls
2. MLP and Backpropagation

N.NN

Algo