SMAI-M20-L22: Introduction to SVM

C. V. Jawahar

IIIT Hyderabad

September 30, 2020

Announcements

- **Quiz 1**
  - The same time as last week.
  - The same set of topics we planned.
  - All objective. (similar to Class Review)

Class Review

In the context of binary classification and LDA:

- What do we know about the direction of the discriminant vector?
- What do we know about the objective?
- What do we know bout the uniqueness of the solution?
- What do we know about the ranks of S_w and S_B?
- How does singularity of S_B and S_w affect us?

obj $w^T A w$

Const $\|w\| = 1$ / or. 2.

Fixed

$w^T S_W w = 1/2$

α^w

w

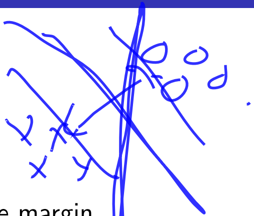
max

min

$\boxed{w^T S_B w \quad w^T S_W w = 1}$

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
    - Kernel Functions and Kernel Matrix



1. Introduction to SVMs
   - SVM as a classifier that maximizes the margin.
2. Solving Logistic Regression as GD
   - From objective to GD and Regularization
3. LDA: Extending to Multi-Class

$\min \ w^T w$

s.t. $y_i(w^T x + b) \geq 1$

Questions? Comments?

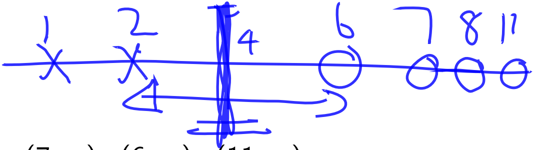
$\boxed{\mathrm{sig}(w^T \cdot x_i) \sim y_i}$

nothing in Am.

Blank

Blank

Consider Five 1D samples:

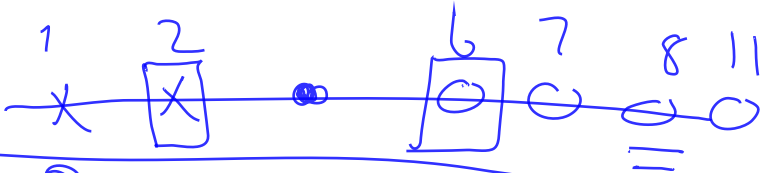


$$(1,+),\ (2,+),\ (7,-),\ (6,-),\ (11,-)$$

1. Which will be a valid decision boundary with perceptron criteria? (i) 2.5 (ii) 6.5 (ii) 5.00 (iv) 4.00
2. What will be the optimal decision boundary with SVM criteria? (i) 2.5 (ii) 6.5 (ii) 5.00 (iv) 4.00
3. Assume we add a sample $(8,-)$ to the training set, will SVM decision boundary change? why? what is the new one?
4. Assume we add a sample $(4,-)$ to the training set, will SVM decision boundary change? why? what is the new one?

Blank

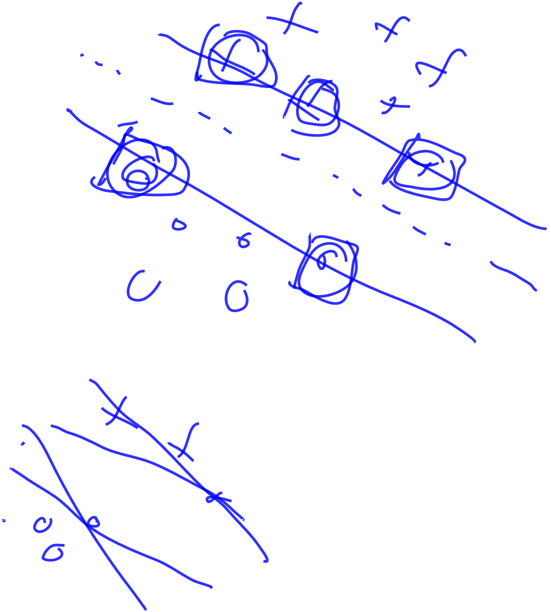
Support vectors



$$w^* = \sum \alpha_i x_i$$

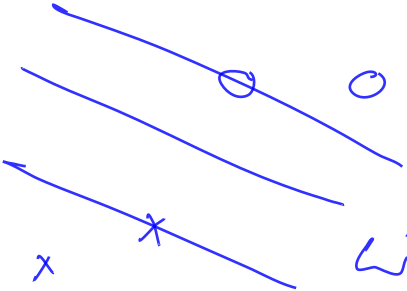
$$w^* = \frac{1}{2}(2) + \frac{1}{t}(6) + 0 \cdot 1 + 0 \cdot 7 + 0 \cdot \ldots \ldots 0 \cdot 11$$

Blank



$wx + b = 0$

Blank



$w^T x + b = -1$

$w^T x + b = 0$

$w^T x + b = +1$

We know the objective of logistic regression as: $y_i = \{0, 1\}$

$$L = \boxed{\sum_{i=1}^{N} y_i \log(g(\mathbf{w}^T\mathbf{x}_i) + (1 - y_i)\log(1 - g(\mathbf{w}^T\mathbf{x}_i)}$$

Derive the gradient ascent update equation

Hint:

$$\boxed{w^{k+1} = w^k + \eta \sum_{i=1}^{N}(y_i - g(\mathbf{w}^T\mathbf{x}_i))\mathbf{x}_i} \leftarrow$$

$$w^{k+1} = w^k + \eta \frac{\partial L}{\partial w}$$

Blank

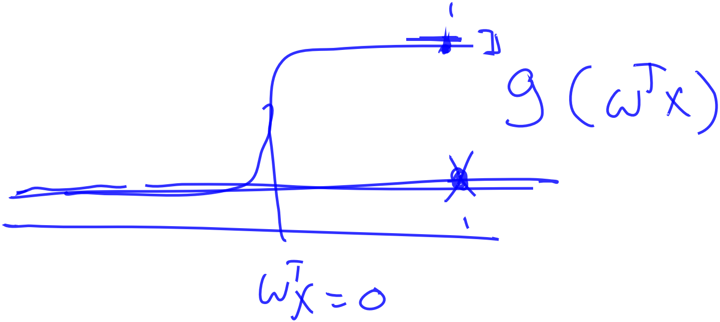
$$\boxed{g'(\theta) = g(\theta)(1-g(\theta))}$$

Blank

Blank

Blank

Blank



$$\sum \left(y_i - g(w^T x_i)\right)^2 \qquad y_i - g(w^T x_i)$$

Discussion Point - III

MLE

X

How do you compare the GD rule with that of if we had used an MSE loss between predicted probabilities and our actual labels (probabilities)?

MLE

$$\boxed{w^{k+1} = w^k + \eta \sum_{i=1}^{N} \underbrace{(y_i - g(\mathbf{w}^T \mathbf{x}_i))}_{\varepsilon_i} \mathbf{x}_i}$$

Blank

Blank

Blank

What Next:? (next three)

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