SMAI-M20-L20: LDA

C. V. Jawahar

IIIT Hyderabad

September 25, 2020

Announcements

1. **Quiz 1** on Next Wed.
2. **Topics:** Topics remain the same.
3. Most-Likely the same time.
4. Any other announcements: by Monday.

Class Review

Consider the following three samples and their labels $((x_1, x_2), y)$:

$$\{((1,1),+),\ \ ((2,2),-),\ \ ((0,0),+)\}$$

Look at the perceptron update rule with $\eta = 0.1$

$$\mathbf{w}^{k+1} \leftarrow \mathbf{w}^k + \eta \sum_{\mathbf{x}_i \in \mathcal{E}} y_i \mathbf{x}_i$$

Classify as + ve if $\mathbf{w}^T\mathbf{x} \geq 0$ else - ve.
Given $\mathbf{w}^0$. What do we know about $\mathbf{w}^1$ and $\mathbf{w}^2$?

Blank

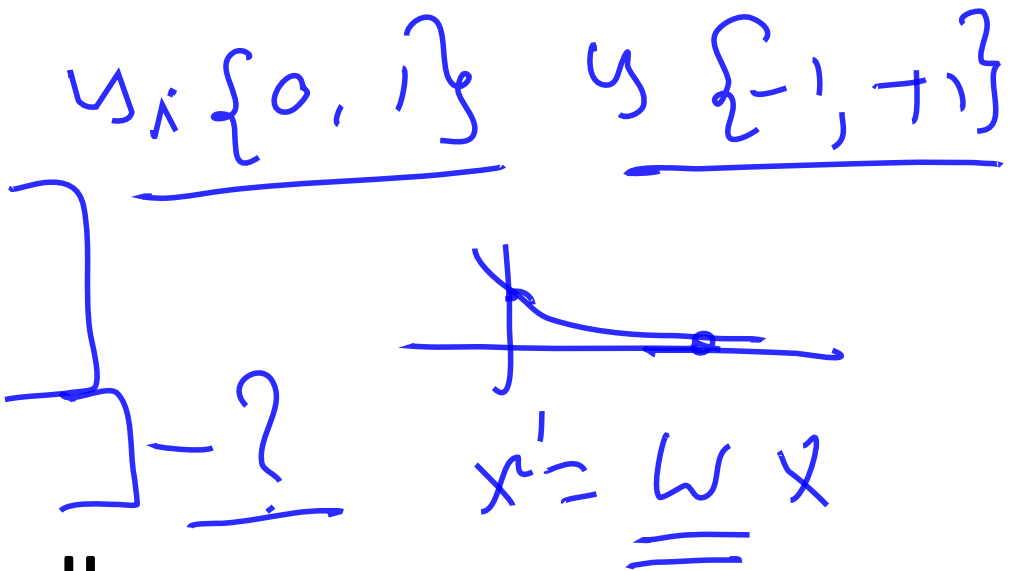
Recap:

- Supervised Learning:
  - Notions of Training, Validation and Testing; Loss Function and Optimization, Generalization, Overfitting, Occam's razor, Model Complexity, Bias and Variance, Regularization.
  - Performance Metrics, Estimating error using validation set.
- Approaches:
  - Optimal Decision as ω_1 if $P(\omega_1|\mathbf{x}) \geq P(\omega_2|\mathbf{x})$ else ω_2, MLE
  - Dimesnionality Reduction and Representation (Feature Selection, PCA, Neural Embeddings)
  - Application of PCA: Eigen Face
  - Matrix Factorization for Data Matrices (SVD, Eigen Decomposition)
  - Application of Matrix Factorization: LSI, Matrix Completion, Recommendation Systems)
  - Nearest Neighbour, Linear Discriminants
  - Gradient Descent
  - Linear Regression: Closed form, GD, Regularization, Optimization
  - Perceptron Algorithm and Neuron Model
  - Logistic Regression
  - LDA
  - Multi-Class Classification Architectures

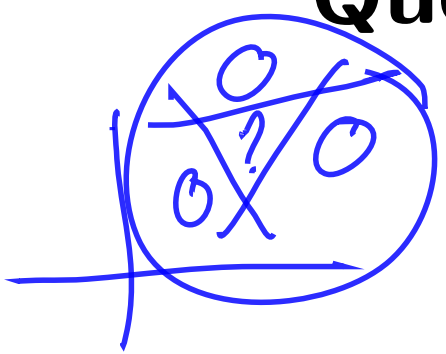
Fisher Face

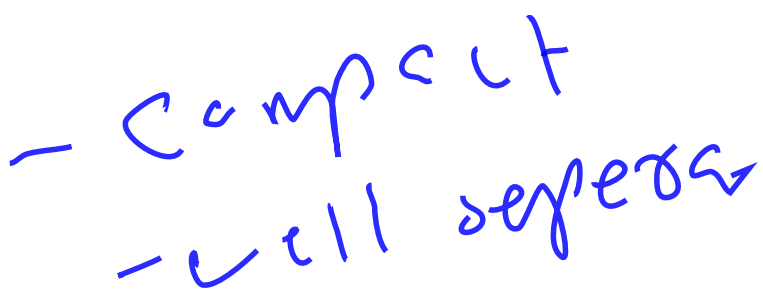
This Lecture:

1. **Logistic Regression - III**
   - Insight into LR objective
2. **LDA - II**
   - LDA solution
3. **Multi-Class Classification - II**
   - D-DAG

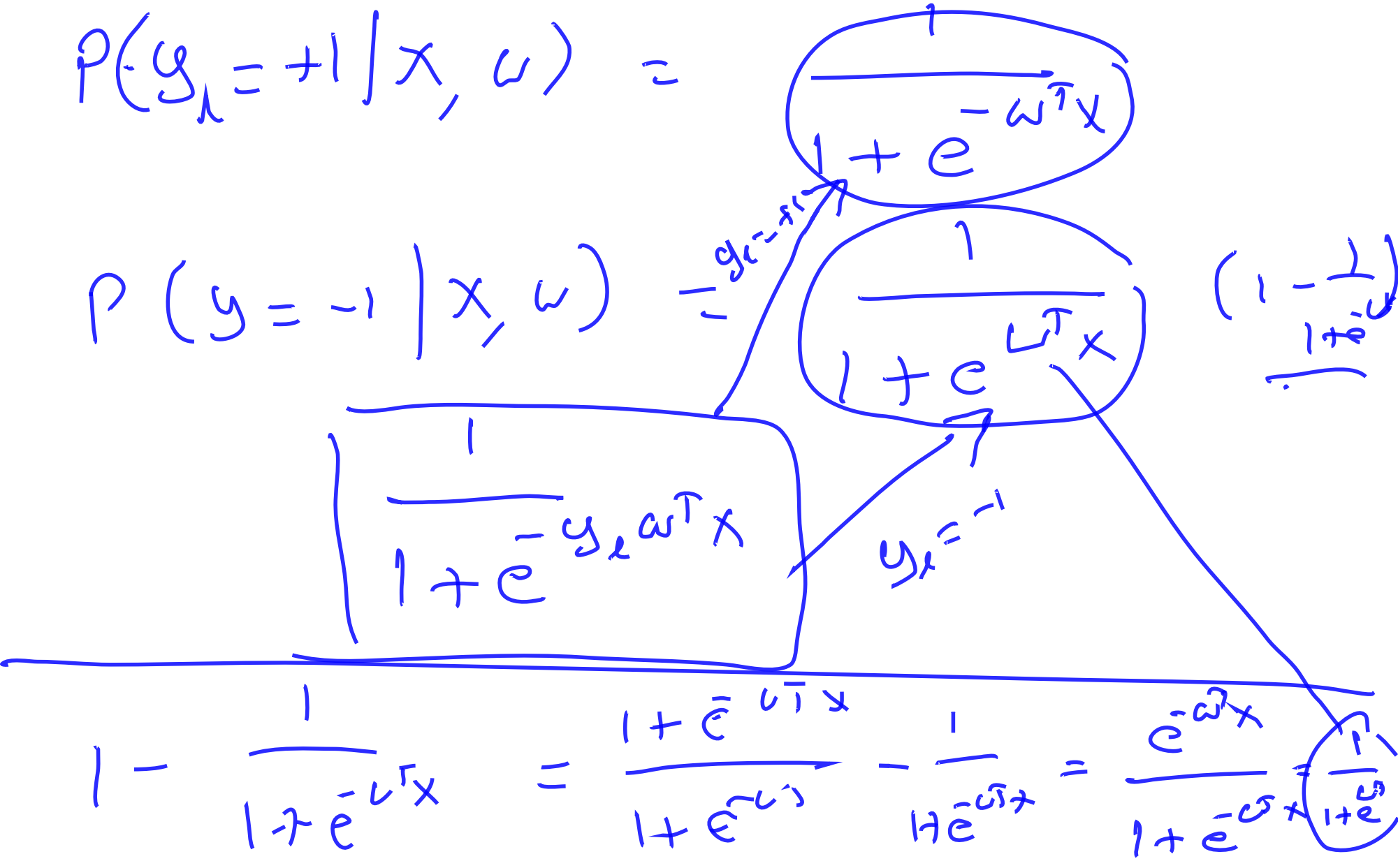


Questions? Comments?





Blank


$P(y_i = +1 \mid x, w) = \frac{1}{1 + e^{-w^T x}}$
$P(y = -1 \mid x, w) = \frac{1}{1 + e^{w^T x}}$
$\frac{1}{1 + e^{-y_i w^T x}}$
$y_i = -1$
$1 - \frac{1}{1 + e^{-w^T x}} = \frac{1 + e^{-w^T x}}{1 + e^{-w^T x}} - \frac{1}{1 + e^{-w^T x}} = \frac{e^{-w^T x}}{1 + e^{-w^T x}} = \frac{1}{1 + e^{w^T x}}$

Discussions Point - I

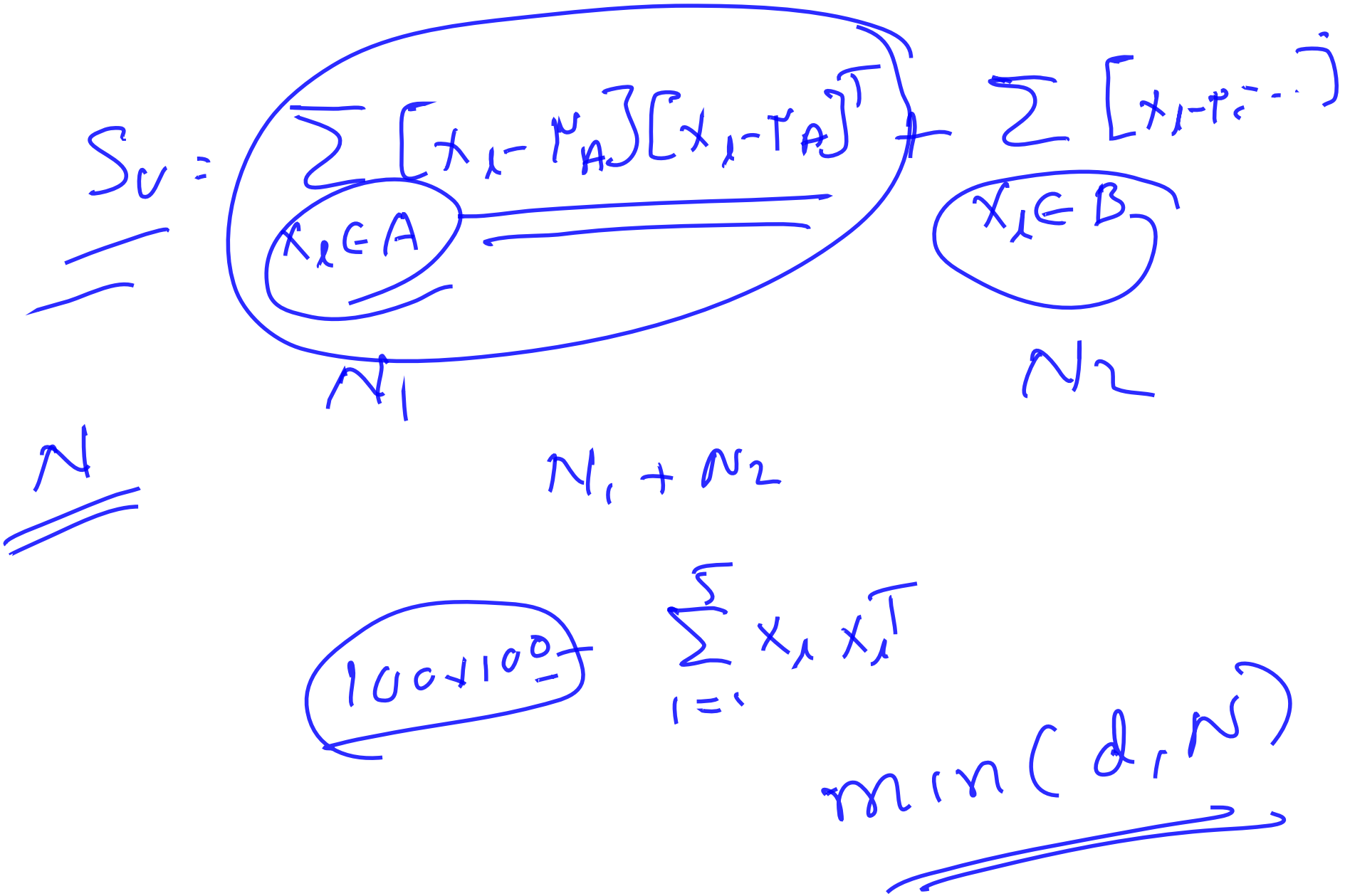
We know the solution to LDA as

$$\mathbf{w}^* = \alpha \mathbf{S}_W^{-1}[\mu_\mathbf{A} - \mu_\mathbf{B}]$$

A potential worry is "If $\mathbf{S_w}$ is singular? "

- Suggest a configuration of the data when $\mathbf{S}_w$ can be singular?
- Suggest solutions to handling this singularity problem while computing $\mathbf{w}^*$?

Blank


$S_U = \sum_{x_i \in A} [x_i - \mu_A][x_i - \mu_A]^T + \sum_{x_i \in B} [x_i - \mu_B][x_i - \mu_B]^T$
N_1
N_2
N
$N_1 + N_2$
100x100
$\sum_{i=1}^{5} x_i x_i^T$
$\min(d, N)$

Blank

① $S_W \leftarrow [S_W + \alpha I]$ (small)

② Pseudo inverse. $S_W^{-1} = (U D V^T)^{-1} = V^{-T} D^{-1} U^T$

$S_W^{-1} = V D^{-1} U^T$

③ Step1 Do PCA $1000 \rightarrow 200$ (d)

$d = 1000$, $N = 500$

Step2 Do LDA S_W $d \times d$ 200×200

Second ω

$$S_B\, \omega = \lambda S_U\, \omega$$

$$[r_A - r_B][r_A - r_D]^T \omega = \lambda S_\omega\, \omega$$

$\alpha = \text{scalar}/\lambda$

$$w = \alpha\, S_U^{-1} [r_A - r_D]$$

$$v = S_U^{-1} [r_A - r_D]$$

$\|v\| = 1$

Blank

"generalized E.V problem"

PCA

$Ax = \lambda x \rightarrow$ E.V Pr

$Ax \quad \lambda Bx \rightarrow$ G. E.V prc

LDA

$$S_B = [r_{A\cdot} - r_{\emptyset}][r_A - r_{\emptyset}]^T$$

S_B is singular

Discussions Point -II

Are there any design considerations in D-DAG?

We know the DAG way of arranging pair-wise classifiers.

(Assume we have 4 classes and 6 pairwise classifiers. We had seen in the micro-lecture.)

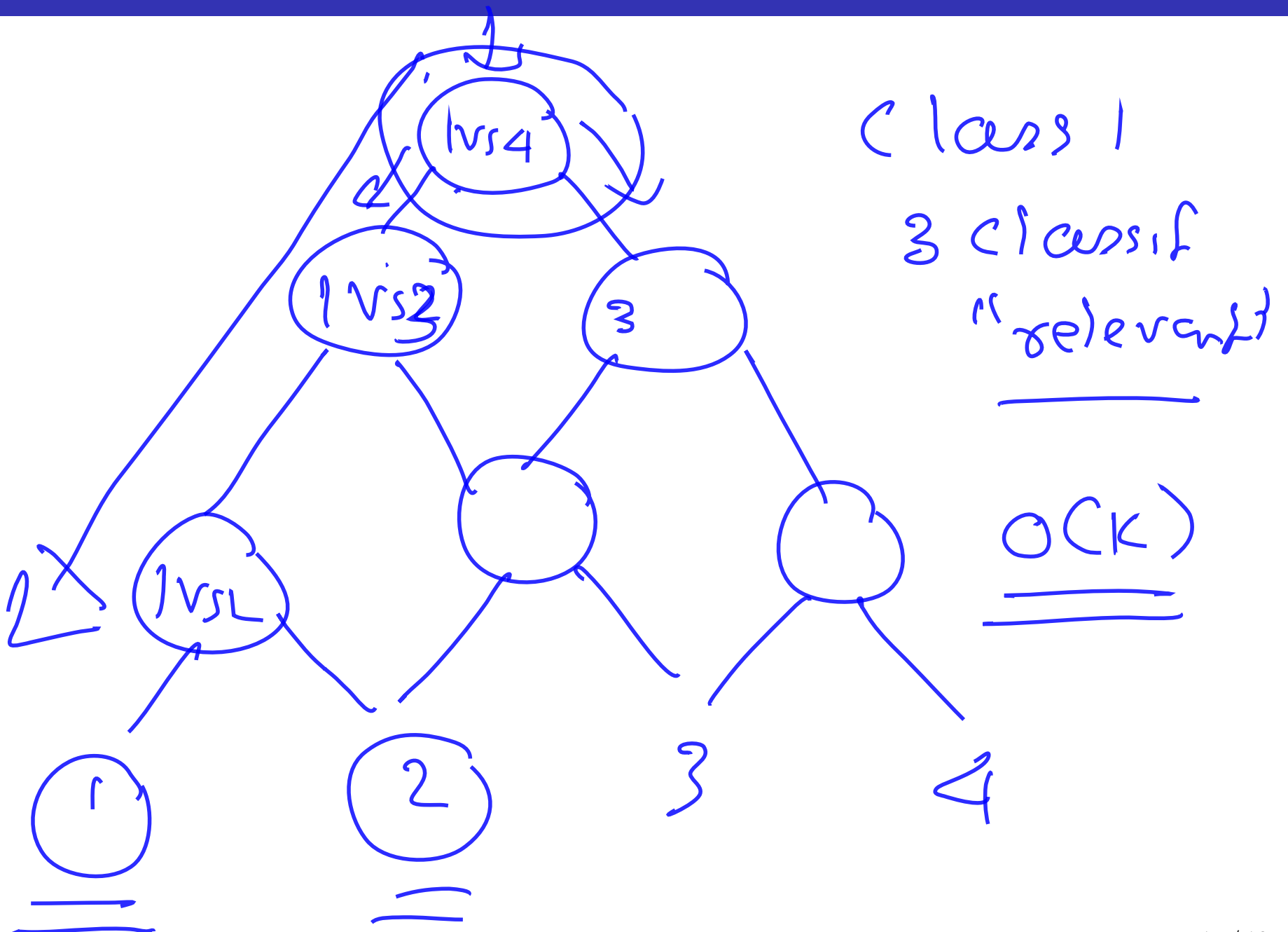
What will you prefer as the root classifier?

- the highest accuracy classifier
- the least accuracy classifier
- any classifier. This does not matter in the design.

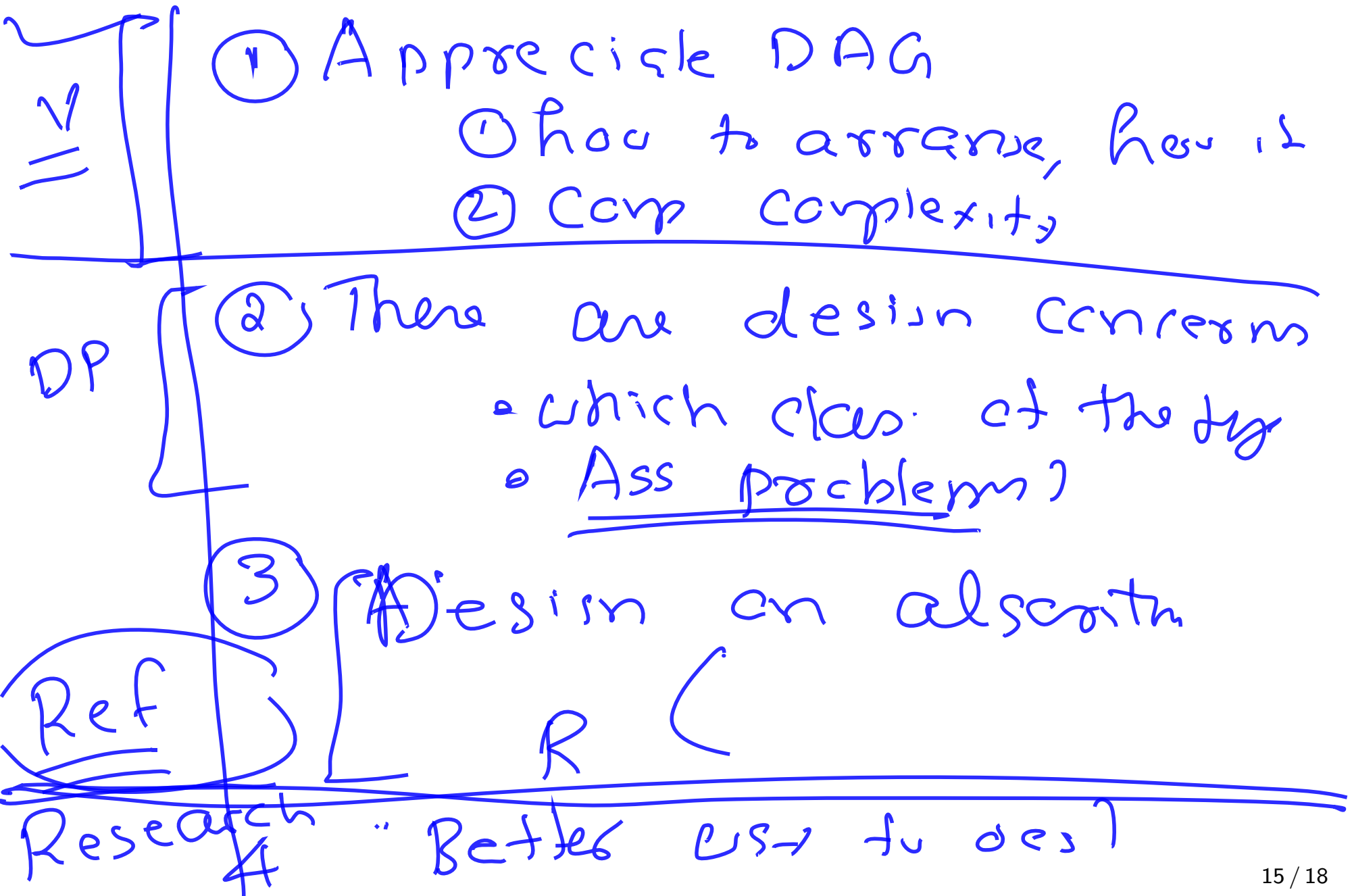
Any other insight into how the classifiers/class to be arranged?[1]

[1] Read later: An old but relevant analysis:
https://cvit.iiit.ac.in/images/ConferencePapers/2003/pavan03multiclass.pdf

Blank



Blank


1 Appreciate DAG
1 how to arrange, how is
2 Comp complexity
2 There are design concerns
• which class of the dp
• Ass problem?
3 Design an algorith
DP
Ref
Research
Better use to des

Discussion Point - III

Comment on the following three different ways of formulating the loss for a binary classification[2]:

1. $\sum_{i=1}^{N}(y_i - \mathbf{w}^T\mathbf{x}_i)^2$
2. $\sum_{i=1}^{N}(y_i - g(\mathbf{w}^T\mathbf{x}_i))^2$
3. LR objective

MSE

why LR did not use this Loss

- Non-convex

MLE

?

[2]Read later: somewhat relevant reference: http://books.jackon.me/Cross-Entropy-vs-Squared-Error-Training-a-Theoretical-and-Experimental-Comparison.pdf

Blank

What Next:? (next three or even more)

1. More on LR, Multi-Class Classification, Dimensionality Reduction
2. Intro to SVMs and Kernels.