# SMAI-M20-L18:Roundup Session

C. V. Jawahar

IIIT Hyderabad

September 21, 2020

# Class Review

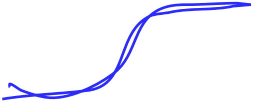
$[0,1]$ $[-1,1]$

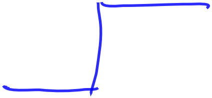
$$\frac{1}{1+e^{-z}} \times \frac{e^z}{e^z}$$

Consider the sigmoid function $g(\alpha z) = \frac{1}{1+e^{-\alpha z}}$

- When $\alpha$ varies what happens to this function?
- Where is its range, what is its value at zero, max value, min value,
- How is this related to tanh? $\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$
- What is the derivative of sigmoid?
- What is $1 - g(z)$?

Sigm
tanh





# 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 $\omega_1$ if $P(\omega_1|\mathbf{x}) \geq P(\omega_2|\mathbf{x})$ else $\omega_2$, MLE
  - Dimesnionality Reduction and Representation ( Feature Selection, PCA, Neural Embeddings)
  - Application of PCA: Eigen Face
  - Matrix Factorization for Data Matrices (SVD, Eigen Docomposition)
  - Application of Matrix Factorization: LSI, Matrix Completion, Recommendation Systems)
  - Nearest Neighbour, Linear Discriminants
  - Gradient Descent
  - Linear Regression: Closed form, GD, RegularizationOptimization
  - Perceptron Algorithm and Neuron Model
  - Logistic Regression

# This Lecture:

6.30 - 7.30

1. Plans and Preparation for the Quiz
2. Topic 1:
   - Feature vector, Data Matrix, Bayesian Optimal Classifier,MLE etc.
   - Eigen Decomposition, SVD, LSI etc.
3. Topic 2: Supervised Learning
   - Training, Testing, Validation, Performance Metrics, Overfitting, Regularization etc.
   - Loss Functions, Optimization, Bias and Variance
4. Topic 3: Algorithms
   - PCA, Linear Regression, Nearest Neighbour, Naive Bayes etc.
   - Perceptrons, Gradient Descent, Logistic Regression (not for Quiz)

**Questions? Comments?**

# Quiz Preparation

- Go through the class material well enough
  - Micro-Lecture Videos
  - Lecture Session (slides/recordings) and some references there in.
  - Class Reviews
  - Home Works
- Read the text book related chapters, Solve problems and practice
- Not expected for Quiz1:
  - Additional references/material list available (moodle)
  - Lecture notes and additional references
  - (https://www.dropbox.com/sh/h91lhc0xpmh2ekw/AAC-FuNgqOO0-txb3FvJqns-a?dl=0)
- Use computer; Keep phone next to you.
- Keep pen, paper and a calculator for your use.
- Strictly avoid any communication with classmates during this period.

# Quiz Questions

**Example 5 Questions; Need not be the same**

1. Test whether you are a student of SMAI 2020 (like CAPTCHA) (may be 10 questions that you can solve fast if you are a student in SMAI-2020) [1]
2. Test whether you not only enrol but also attempt to follow[2] SMAI-2020.
3. Numerical Problems
4. A question set very similar to the regular class review one.
5. A bit more involved. Keep couple of A4 sheets and a paper ready.

[1] CAPTCHA stands for Completely Automated Public Turing test to tell Computers and Humans Apart.

[2] Follow Class Material

# Topic - I

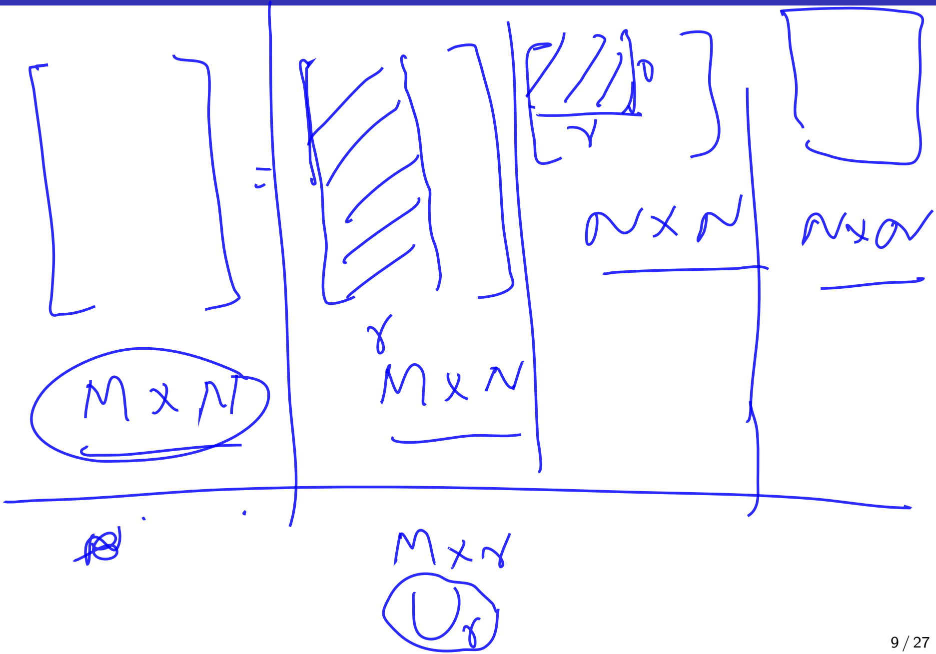
$\{(c_1, c_2))$

- Feature vector, Data Matrix, Bayesian Optimal Classifier,MLE etc.
- Eigen Decomposition, SVD, LSI etc.
- Q: In posterior probability P(w1/X), does X denote class label or feature in the vector x?

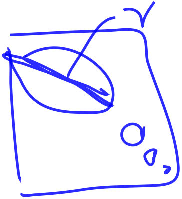
$P(w|x)$ $P(x|w)$

- Yes;
- In SVD we do $A = UDV^T$. The dimensions of U are given as m x r, but we know that U is the matrix of eigen vectors of $AA^T$, and $AA^T$ is symmetric so it is supposed to have independent eigen vectors. If we assume a to be M x N matrix, the dimension of you must be M X M not M X R, sinec all the eigen vectors are linearly independent.

# Blank



# Blank



m > n

n > m

m x n

~~rank - r~~

$$\underset{n \times n}{A} = \sum_{i=1}^{r} \underset{n \times 1}{x} \, x'$$

$m > n$
$n \geq m$

$$A = U D V^T$$

$$A' = V \underline{\underline{D}} U^T$$

# Blank

# Blank

eigenvect

PSD

- Training, Testing, Validation, Performance Metrics, Overfitting, Regularization etc.
- Loss Functions, Optimization, Bias and Variance
- Q: Given a regularization problem, how do we minimise the loss?

$$L = \text{EmpL} + \text{RegTerm}$$

$$L = \sum_{i=1}^{N} \ell(y_i, f(x_i)) + \|w\|_2^2$$

# Blank

$$L = L_1 + \lambda L_2$$

$$\boxed{\frac{\partial L}{\partial w} = \frac{\partial L_1}{\partial w} + \lambda \frac{\partial L_2}{\partial w}}$$

# Blank

$$L = \min_w \boxed{\sum (y_i - w^T x_i)^2} \quad =0$$

$$\max_w \boxed{w^T A w} \qquad 2Av$$

$$w^T v = 1$$

$$w^0$$

$$\boxed{w^{k+1} \leftarrow w^k - \eta \nabla L}$$

# Blank

$L \quad w^T A w$ $\qquad \|w\| = 1$

$\frac{\partial L}{\partial w} = 2Aw$

$w^0 \leftarrow \text{random}$

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

② normalize $w^{k+1}$ s.t. $\|w^{k+1}\| = 1$

# Blank

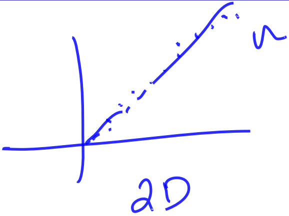
# Topic - III

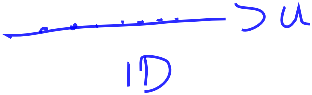
- PCA, Linear Regression, Nearest Neighbour, Naive Bayes etc.
- Perceptrons, Gradient Descent, Logistic Regression (not for Quiz)

# Blank

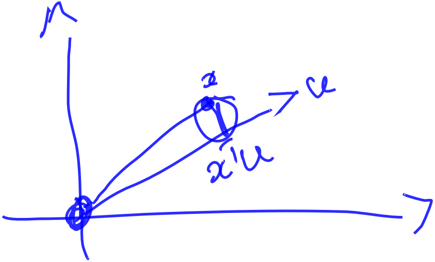
$x_i \to x_i^T u$

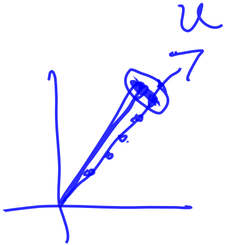
$\alpha_1$




- Q: Steps of minimizing loss functions to obtain $\min ||X - X'||$



# Blank

Let $\mathbf{u_1}, \dots \mathbf{u_d}$ be $d$ orthonormal vectors. We can represent the vectors $\mathbf{x}$ as $\sum_{i=1}^{d} \alpha_i \mathbf{u}_i$. Where the scalar $\alpha_i$ is $\mathbf{x}^T\mathbf{u_i}$. However, if we use smaller than $d$ basis vectors, there could be some loss or reconstruction error. Let us consider the loss when we use only one $\mathbf{u}$. i.e.,

$$\mathbf{x} - \mathbf{u}\mathbf{u}^T\mathbf{x} \quad \alpha \qquad — \ 1$$

Sum of the reconstruction loss for all the $N$ data samples is now:

$$\boxed{\sum_{i=1}^{N} ||\mathbf{x_i} - \mathbf{u}\mathbf{u}^T\mathbf{x}_i||^2} \qquad u^T u = 1 \quad 2$$

$$= \sum_{i=1}^{N} \left( \mathbf{x}_i^T\mathbf{x}_i + (\mathbf{u}\mathbf{u}^T\mathbf{x}_i)^T(\mathbf{u}\mathbf{u}^T\mathbf{x}_i) - 2\mathbf{x}_i^T\mathbf{u}\mathbf{u}^T\mathbf{x}_i \right) \qquad 3$$

We would like to minimize this. First term is positive (non negative). It is independnet of $\mathbf{u}$. Therefore, we would like to minimize:

$$\sum_{i=1}^{N} (\mathbf{x_i}^T\mathbf{u}\mathbf{u}^T\mathbf{u}\mathbf{u}^T\mathbf{x_i} - 2\mathbf{x}_i^T\mathbf{u}\mathbf{u}^T\mathbf{x}_i) \qquad 4$$

We also know that $\mathbf{u}^T\mathbf{u} = 1$.

$\min \| x - x' \|$

$$= \sum_{i=1}^{N} -\mathbf{x}_i^T \mathbf{u}\mathbf{u}^T \mathbf{x}_i = \sum_{i=1}^{N} -\mathbf{u}^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{u} = -\mathbf{u}^T \Sigma \mathbf{u}$$

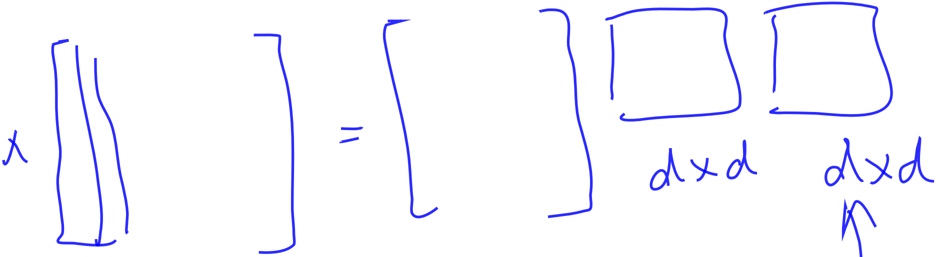
Minimizing the reconstruction error now becomes that of Maximizing

$$\text{Max} \quad \mathbf{u}^T \Sigma \mathbf{u}$$

with our familiar constraint of $\mathbf{u}^T\mathbf{u} = 1$. This reduces the solution as the eigen vectors corresponding to the largest eigen values.

$$\boxed{\text{Max } u^T \Sigma u \text{ s.t } u^T u = 1}$$

# Blank




N x d

X

N x d

$$\boxed{\frac{1}{N}\sum_{i=1}^{N}[x_i-\mu][x_i-\mu]^T}$$

Eig vec of $xx^T$

Ei of $X^TX$

Covari

# Blank

# Blank

# Blank

# What Next:? (next three)

1. Logistic Regression
2. Multi Class Classification (beyond binary)
3. More Dimensionality Reductioin Schemes (eg. LDA/Fisher)