SMAI Endsems

Mid 1

- Basics
  - Generalised Gradient Descent
  - Learning Rate Optimisation
  - Newton's Updates [Optimised Update Rule]
  - Proof that gradient Descent works
  - Pattern Recognition Flow Diagram (7+5)
  - Gaussian distribution Formula
- Linear Algebra
  - L-Norm
  - Span of set
  - Linear Independence
  - Rank of a matrix
  - Basis, orthonormal Basis
  - Eigen Values
  - Eigen Vectors
- Basics of ML
  - Regression
  - Classification
  - Discriminant Function
  - Probabilistic approach to classification
  - Multi-class Classification Problem
- Regression
  - Basics of Regression
  - Error and Update Functions
  - Derivation of Normal Form
  - Pseudo Inverse of a Matrix and why its Needed
  - PCA shit, and it's derivation of XtX = Lambda
- Linear Regression
  - Cost/Error/Loss Function
  - Hypothesis Function
  - Update Function
  - Learning Rate
  - Learning Rate Update Function
  - Theory
- Logistic Regression

    - Cost/Error/Loss Function
    - Hypothesis Function
    - Why not linear classifiers
- Perceptron
    - Discriminative Classifier
    - Generative Classifier
    - Probability based classification.
    - The Perceptron Algorithm [net, sign, delta omega, omega]
- Bayes' Decision Theory
- Regularisation
- SoftMax

Mid 2

- PCA
    - *Dimensionality Reduction*
        - Feature extraction
        - Feature selection
    - Two points on PCA
        - Dimensionality reduction that retains maximum information
        - Reconstruction of original data with minimum error
    - Proof of mean projection
    - **EIGEN VECTOR AND VALUE**
    - Minimising Reconstruction Loss - Proof
    - Retention of maximum information - Proof
    - **PCA - Algorithm - All Steps and Why**
    - Deciding k
    - Eigen Faces
    - N and d value comparison and inter-change calculations
- SVM
    - Why SVM
    - Constraints for SVM
    - Primal problem equation
    - Dual problem equation
    - Support Vector
    - Number of Support Vectors and Error
    - Soft SVM and new constraint equation
    - Implementation Algorithm
    - Priman to Dual
- Kernels

    - Why Kernel
    - Feature Map
    - Kernel
    - Kernel Matrix and properties
    - Non-linear, Kernel SVM
    - Example Breakdown
    - Popular Kernels
    - Kernel to Feature Map [*Always dot product of feature map*]
- LDA and KPCA
    - LDA, FDA
    - Within Scatter, Between Scatter
    - New Objective Function and Minimisation
    - Generalised Eigen Value problem and Sw inversion
    - Sw defined wrt each component pair
    - KPCA equation
    - KPCA Calculations
    - Projection and Centring of Kernel
    - Steps
- MLE Bias
    - Proof of mean estimator unbiased
    - Proof of variance estimator biased
    - Derivation of variance estimator modification for unbiase

Unsupervised Learning - Notes

    - Clustering:
      Broadly tree requirements.
      - Proximity Measure
      - Criterion Function
      - Algorithm
    - K-Means
    - Distance Calculation
        - Euclidean
        - Manhattan
        - Minkowski
        - Mahalanobis Distance
    - Gaussian Mixture Model
    - Soft Clustering
    - E-M Algorithm
    - Cluster Validity
        - Stability

        - Average Distance to Center
    - Hierarchical Clustering
        - Top Down Clustering
        - Bottom Up Clustering
        - Single Link Clustering Algorithm
        - Complete Link Clustering
    - Measuring How Good is Clustering:
        - Purity Metric
        - Rand Index and Table 2x2
    - Semi-supervised Learning
    - Bayes Rule
    - Probabilistic Graphical Model
    - Conditional Independence
    - Inter Casual
    - Hidden Markov Model
    - Variable Scaling and it's importance

http://www.mit.edu/~9.54/fall14/slides/Class13.pdf
https://lagunita.stanford.edu/c4x/HumanitiesScience/StatLearning/asset/unsupervised.pdf
https://www.andrew.cmu.edu/user/achoulde/95791/lectures/lecture07/lecture07_95791.pdf
https://www.youtube.com/watch?v=REypj2sy_5U

- Cluster evaluation:
    - Intra Cluster Cohesion
    - Inter Cluster Separation
- Hierarchical Algorithms - Using previously established clusters
    - Agglomerative : Bottom-up
    - Divisive : Top Down
- Partitional - Determines all clusters at once.
- Bayesian - Posteriori Distribution
    - Distance Calculation
        - Euclidean
        - Manhattan
        - Minkowski
        - Mahalanobis Distance

- K-Means (*A partitional clustering Algorithm*)
    - The Algorithm.
    - Stopping Point (3)

    - Strengths:
        - Easy.
        - O(tkn)
    - Weaknesses:
        - Requires Mean
        - Specification of k
        - Sensitive to outliers
        - Sensitive to Seeds
        - Special Data Structures
    - Dealing with Outliers:
        - Remove datapoint
        - Random Sampling
    - Proof of Convergence

- Hierarchical Clustering
    - Dendrogram
    - Divisive
        - Recursive K-means with k = 2
    - Agglomerate
        - Single Linkage : NN - MST
        - Complete Linkage : FN
    - Divisive vs Agglomerative
    - Linkages:
        - Single
        - Complete
        - Average
        - Centroid

- Gaussian Mixture Model - EM Clustering
    - Intuition
    - Algorithm
    - Steps:
        - Expectation
        - Maximisation
    - Perks
    - Updation through responsibility calculation
        - mu updation formula
        - Sigma updation formula

Supervised Learning - Lesson 17,18,19

- Single layer perceptron and Activation function
- Architecture and Non-Linearity
- Loss Functions :
  - MSE
  - Cross Entropy
  - Regularised Loss
- Activation Functions:
  - Sigmoid
  - Tanh
  - ReLu
  - Leaky RelU
- Implementations for OR/NOR/AND/NAND
- Implementation for EXOR
- Derivatives and updation in MLP
- Loss Layer Calculation in MLP
- Fully Connected Layers
- Forward Propagation
- Backward Propagation
  - Basic Intuition and Math
  - Algorithm (5)
  - Optimisations (4)
  - Stochastic Mini-Batch
  - Epoch
  - Sub-gradients
- Xavier's Initialisation
- Nesterov Momentum
- ADAM, AdaGrad, AdaDelta, RMSProp
- Termination Criteria
- Softmax
- KL Divergence
- Regularisation
- Hessian Methods (**Newton's Method**)
- Vanishing and Exploding Gradients

Supervised Learning - Notes

- Decision Trees
- Entropy
- Gini Impurity
  - Advantages
  - Limitations

- Overcoming Overfitting
  - Early Stop
  - Pruning
- Dealing with missing data
- Rain Forest Classifier
  - Majority Vote
  - Weighted Voting
- Feature Level Fusion
- Bagging - Bootstrap Aggregation
- Boosting - adding weights to samples
- Weakening Decision Trees:
  - Ignoring Features
  - Stumping
  - Bagging
  - 200 Averaging

Semi-Supervised Learning - Notes

- Constraints and Softening
- Probabilistic Graphical Model
- Hidden Markov Model