Second Semester Area B (data science)

CSE 382M, Foundational Techniques of Machine Learning and Data Sciences

1. **Introduction to Data Science:** Historical perspective, Applications.

2. **High-Dimensional Probability:** Multivariate Gaussians, CLT, Concentration of measure.

3. **Linear Algebra:** SVD, Condition number, Overdetermined / underdetermined systems.

4. **Regression / Interpolation:** Generalization/overfitting, Regularization, Cross-validation.

5. **Optimization:** Gradient Descent, Convexity, Smoothness.

6. **Stochastic Optimization:** Stochastic Gradient descent, Importance sampling, Condition number, Implicit regularization.

7. **Classification:** Support Vector Machines (SVM), Kernel SVM Classification

8. **Principal Component Analysis (PCA):** Best rank-k approximation, Power method.

9. **Random Projections:** JL lemma, Approximate K-nearest neighbor

10. **Matrix Sampling and Sketching.**

11. **Compressed Sensing:** Sparsity, L1 regularization.

12. **Learning Models from Data:** Convex versus non convex optimization, Alternating minimization / expectation maximization.

13. **Unsupervised Clustering:** k-means, Lloyd's method, Spectral clustering

14. **Statistical Estimation:** Bayes rule , Maximum likelihood, Maximum a posteriori (MAP) probability

15. **Graph Analysis:** Random walks, Markov chains, Community detection.

16. **Control Theory:** Hidden Markov models, Kalman filter / Linear quadratic estimation

17. **Neural Networks:** Resnet for supervised learning, Generative adversarial network for unsupervised learning, Implicit regularization, Overparameterization.