

Plan:

Feb 3 includes the motivation for data science, mathematical foundations-Lagrangians, data visualization, scatter plot, least squares, regression models and basics of R.

Feb 10 includes an introduction to hypothesis testing, design of experiments, factor analysis and applications in CSPA, confidence intervals.

Feb 17 covers talks on applications in several areas in Science and Engineering.

*****Relevant Data Science Questions*****

- Q1: What steps do you take after the data are in hand?
 - Q2: Which are your pet type of projects (e.g., multi-classification).
 - Q3: When is R squared meaningless?
 - Q4: When is RMSE more meaningful than R squared?
 - Q5: Why and how do you straighten data?
 - Q6: How do logistic regression (LR) and probit regression differ?
 - Q7: What rule, i.e., 80/20, do you use to split the data for model validation?
 - Q8: How do you convert logits into probabilities?
 - Q9: What's wrong with using OLS regression with a binary DepVar?
 - Q10: Which fitness measures do you use for a given model (i.e., clustering)?
 - Q11: Define Cum Lift, aka, Up Lift?
 - Q12: How do you rank the importance of predictors in a regression model?
 - Q13: Standardized reg. coeffs. are useful with OLS. Are they useful with LR?
 - Q14: Compare Factor Analysis (FA) and Principal Component Analysis?
 - Q15: Compare FA and Latent Class Analysis?
 - Q16: Compare FA and little-known Q Analysis?
 - Q17: How do you handle the mess that missing data make?
 - Q18: Define the Chance Model, and why should you always use it?
 - Q19: How do you remove noise from data?
-

Booklist:

- The Elements of Statistical Learning
- Reinforcement Learning: An Introduction
- Applied Predictive Modeling
- Pattern Recognition And Machine Learning
- Computer Age Statistical Inference
- Mining of Massive Datasets

Course list:

<https://www.amazon.in/Python-Data-Science-Handbook-Essential/dp/9352134915>

- Topics in Mathematics with Applications in Finance (MIT): <https://lnkd.in/gU4bfsq>
- FAST.ai part2: <https://lnkd.in/gEvbYp4>
- CS 294: Deep Reinforcement Learning: <https://lnkd.in/gU4bfsq>
- CMU 701 by Tom Mitchell: <https://lnkd.in/gU4bfsq>
- Statistical Rethinking: <https://lnkd.in/gUWjsYv>
- Probabilistic Graphical Models: <https://lnkd.in/gvz59AG>
- NTU - Machine Learning (2017,Fall): <https://lnkd.in/gYeuzsb>