Exam 1 Material:

1. Why is ML an ill-posed problem - consistency and generalizability
2. Inductive bias (understand restriction bias vs preference bias) - for ex: multivar linear reg or decision tree
3. Sampling bias and potential issues (think of train vs deployment)
4. What is meant by no free lunch theorem in ML?
5. Model complexity vs simplicity (bias vs variance tradeoff, underfit vs overfit and why) - how to handle high variance vs how to handle high bias?
6. Raw features vs derived features
7. Interpret a data quality report and identify data quality issues. What to analyze for continuous vs categorical features?
8. Distributions of data (uniform, normal, exponential, skewed, multimodal, etc..)
9. Imputation (continuous vs categorical) and handling outliers
10. Interpret histograms, bar plots, multiple bar plots, stacked bar plots, box and whiskers, scatter plots, correlations
11. Normalization, binning, and sampling techniques - and when / how to apply?
12. Relation between entropy and probability
13. Information gain: initial entropy vs remainder entropy, and how it is applied in DT to select nodes
14. How DT is built, from root to leaf, and interpret final tree (e.g. fundamental understanding of how tree on Slide 4A pg 58 or 4B pg 28 is constructed / interpreted)
15. How continuous features and targets are handled in DT
16. DT pruning - why prune and how pre pruning (early stopping) vs post pruning differ (feature interaction effects?)
17. Why model ensembles?
18. How does Random Forest work vs boosting vs gradient boosting? What are decision stumps?
19. Pros and Cons of DT models
20. How to train a KNN? How KNN works? How to retrain with new data?
21. Concept of eager learner vs lazy learner in ML (KNN vs DT vs Lin Reg vs NB…etc)
22. How does chosen k-value impact results of KNN (in terms of overfitting and underfitting)? Risks with imbalanced data sets?
23. How does distance weighted KNN improve traditional KNN (what does it do? why it potentially improves outputs in some situations?) K sensitivity of weighted knn?
24. Why does data normalization matter (in general, and specifically for KNN)?
25. KNN with continuous targets
26. Curse of dimensionality (why is there a tradeoff between # of features and density in feature space?). How does it impact models (e.g. DT vs KNN vs Lin Reg vs NB…etc)?
27. How to restrict # of features to a subset that is most useful (forward vs backward selection)?
28. Be able to calculate probabilities, conditional probabilities, joint probabilities from a simple table (e.g. slide 6A pg #26)

29. Challenges with predicting rare events from a Bayesian perspective (paradox of false positives)
30. What makes Naive Bayes Classifier, naive? (i.e. understand conditional independence between vars). Even though naive, why does it still perform well?
31. Toy problem of NB classifier similar to slide 6A pg # 60-65
32. Regression (interpret equation) and Loss Function (explain which model is better in terms of residuals, sum squared error, location on error surface)
33. Gradient descent: how an error surface is generated from weight space, why convexity is important, how algo ‘descends’ by adjusting weights (how are weights updated), minimum points
34. Impact of learning rate on gradient descent (in terms of speed, convergence); how learning rate schedulers (decay) work and impact on gradient descent speed, convergence
35. Impact of outliers on linear regression
36. Categorical features in regression - ordinal encoding vs dummy variables
37. Logistic regression: when to use? how the thresholding [Logistic(w·d)] works. How normalization helps with sensitivity to learning rate / initial weight params.
38. How to handle non linearity? What do basis functions introduce (how do they transform raw inputs but keep the model as linear combination of weights)?
39. How does one vs all multinomial logistic regression work? (functionally, at a high-level)
40. Describe an SVC (SVM Classifier) in terms of how it selects the decision boundary (separating hyperplane)
41. Fundamentally, what does a Kernel transformation do (e.g. x^2 example in class)?
42. What are the objectives of regularization? (e.g. ridge and lasso regression)
43. In general, be able to compare / contrast different models we have discussed (pros, cons, weaknesses, their applications, etc)