# Concepts to know

Bias variance tradeoff

bias is error in a given model, variance is the measure of variability between the bias on different sets (ie the train and test sets). High variance can mean overfitting, high bias can mean underfitting.

ROC curve

Graphical representation of contrast between true positive rates and the false positive rate at various thresholds. Helps us see the tradeoff for sensitivity vs false positives with different parameters.

Precision and recall

Recall is the true positive rate which is the number of number of predicted positives divided by the total number of positives in the dataset. Precision is the number of accurate positives the model predicts divided by the number of positives it predicted.

### REVIEW THIS

Bayes theorem

Bayes theorem is conditional probabilities. It is the true positive rate of a condition divided by the sum of the false positive rate of the population and the true positive rate of a condition.

Type 1 and type 2 error

Type 1 error is a false positive and type 2 error is a false negative.

Probability vs likelihood

Probability is the integral over the probability distribution. Likelihood is the likelihood that a value will be exactly some value. It is the corresponding point on the probability curve to a variable.

Generative vs discriminative model

Generative learns categories or things about a dataset while a discriminator learn the differences between the categories of data.

Cross-validation on time series

Forward chaining is the best way to do this. It is using training [1] test [2] then training [1 2] test [3] and so on.

Decision tree pruning

Branches with weak predictive power are removed to reduce complexity. Reduced error pruning is removing nodes one by one, if it doesn’t decrease predictive accuracy keep it pruned

Performance vs accuracy

Accuracy is the score of a model on a given dataset, it may be skewed by class imbalance or other things. Performance is the more abstract concept of how good a model is at doing its intended job.

F1 score

Weighted average of precision and reacall. Useful when true negatives don’t matter much.

Imbalanced dataset

Decision trees are good on imbalanced datasets, it just tries to separate classes. We could also add class weighting to change the error on different classes.