

| | | |
|------|---|--|
| 200 | What is the central limit theorem? | it states that if we sample from a population using a sufficiently large sample size, the mean of the samples (also known as the sample population) will be normally distributed (assuming true random sampling). What's especially important is that this will be true regardless of the distribution of the original population. |
| 400 | what is the difference between parametric and non-parametric models? | parametric - you determine coefficients and y intercepts assumes a distribution about the data non-parametric - does not assume a distribution, does not calculate coefficients |
| 600 | what is the difference between a count vectorizer and tfidf vectorizer? | |
| | how does adaboost work? | |
| 800 | name 5 ways to evaluate a classification model. | classification report - look at F1 score (the harmonic mean between precision and recall) accuracy score auc score check out your confusion matrix to calculate TP TN FP FN look at recall score for each class |
| 1000 | What are the 5 assumptions of linear regression? | Linearity: is linearly related to for all . Independence: Independence of errors Normality: The errors (residuals) follow a Normal distribution with mean 0. Equality of Variance: The errors (residuals) should have a roughly consistent pattern, regardless of the value of . (There should be no discernable relationship between and the residuals.) Independence Part II: independence of variables |