## Name :

**NetID**:

## Please specify whether each statement is True or False:

1. One should never use cross-validation if a dataset has more than a million records: \_ \_\_\_\_\_\_\_\_\_
2. Model estimation bias is not independent of sample size: \_\_\_\_\_\_\_\_\_\_
3. The AUC is a metric that is suitable for ranking problems:\_\_\_\_\_\_\_\_\_\_
4. A “hyper-parameter” in a classification model is often a variable that controls the complexity of the underlying model: \_\_\_\_\_\_\_\_
5. The best classification threshold for any probabilistic classifier is 0.5 : \_\_\_\_\_\_\_\_\_
6. When irreducible error exists one can never build a classifier that has perfect out-of-sample predictions: \_\_\_\_\_\_\_\_\_
7. If false negatives are much more expensive than false positives, then one should aim to maximize a classifier’s recall: \_\_\_\_\_\_\_\_\_\_
8. If your modeling goal is to produce well calibrated probability estimate, log-loss would be your best evaluation metric: \_\_\_\_\_\_\_\_\_
9. Given a single dataset, model estimation variance tends to increase as one uses algorithms with increased complexity: \_\_\_\_\_\_\_\_\_\_
10. If any classifier has a 0-1 Loss of 50%, then it is no better than a random classifier: \_\_\_\_\_\_\_\_\_