

Instructions: Answer each question as thoroughly as possible. Round answers to 4 decimal places as needed. Exact answers are best when possible. Be sure to answer all parts of each question.

1. Describe the difference between unsupervised learning and semi-supervised learning.

Unsupervised learning has no labels on the data to compare results to. In semi-supervised learning, labels are available but not used directly in the model. They are used after the fact to validate results.

2. Explain how factor analysis is related to correlation.

factors are created by finding out which variables are most closely associated w/ each other, i.e. are highly correlated.

3. Describe two advantages and at least one disadvantage of neural networks.

They can handle complex, non-linear data

Can make accurate predictions

however, they are a black box and are not interpretable

(answers may vary)

4. Describe leave-one-out cross validation and why we use it to validate classification models.

LOOCV uses all but one observation to build a model

and then predicts the remaining observation as Yes or No
(0 or 1)

each observation takes a turn being left out.

Since all predictions are either 0 or 1, this allows us to use the most data in creating the model and still get a good sense of its accuracy.