

Homework 1

PSTAT 131/231

Nicole Magallanes

Contents

Main Ideas	1
Exploratory Data Analysis	2

Main Ideas

Question 1

Define supervised and unsupervised learning. What are the difference(s) between them?

Answer: Supervised learning is a machine learning approach that is built to train or supervise algorithms to estimate or predict outcomes based on the given inputs. This approach uses labeled inputs and outputs. Unsupervised learning is an approach that takes in inputs but does not have any supervising outputs. These algorithms can help to analyze and cluster unlabeled data. The main difference between these two approaches is that unsupervised learning does not have any supervising outputs, while supervised learning has labeled inputs and outputs. Also, in supervised learning the algorithm learns from its given training set while unsupervised learning mostly works on its own to discover patterns within the unlabeled data. For the most part, supervised learning models tend to be more accurate than unsupervised models.

Question 2

Explain the difference between a regression model and a classification model, specifically in the context of machine learning.

Answer: The difference between a classification and a regression model is that classification models in machine learning are used to assign data into specific categories while regression models are used to understand the relationships between dependent and independent variables. An example of a classification model in the real world can be spam detection in emails, an example of a regression model can be predicting the impact of test scores on college admissions.

Question 3

Name two commonly used metrics for regression ML problems. Name two commonly used metrics for classification ML problems.

Answer:

Question 4

As discussed, statistical models can be used for different purposes. These purposes can generally be classified into the following three categories. Provide a brief description of each.

Answer:

Descriptive models:

Inferential models:

Predictive models:

Question 5

Predictive models are frequently used in machine learning, and they can usually be described as either mechanistic or empirically-driven. Answer the following questions.

Define mechanistic. Define empirically-driven. How do these model types differ? How are they similar?

Answer:

In general, is a mechanistic or empirically-driven model easier to understand? Explain your choice.

Answer:

Describe how the bias-variance trade off is related to the use of mechanistic or empirically-driven models.

Answer:

Question 6

A political candidate's campaign has collected some detailed voter history data from their constituents. The campaign is interested in two questions:

Given a voter's profile/data, how likely is it that they will vote in favor of the candidate?

How would a voter's likelihood of support for the candidate change if they had personal contact with the candidate?

Classify each question as either predictive or inferential. Explain your reasoning for each.

Answer:

Exploratory Data Analysis

Exercise 1

Exercise 2

Exercise 3

Exercise 4

Exercise 5