



Applied Remote Sensing Training (ARSET) Program

Fundamentals of Machine Learning for Earth Science

Homework Questions

Question 1

What are the two most common types of supervised learning?

Answers: (bold correct)

- a. Classification and clustering
- b. Clustering and regression
- c. **Classification and regression**

Question 2

What are some applications of machine learning for Earth Science?

Short Answer:

Question 3

True or False: Physics-informed machine learning models are exclusively data driven.

Answers: (bold correct)

- a. True
- b. **False**

Question 4

In Random Forest or XGBoost (ensemble learning algorithms), one adds predictions for weak learners so that an ensemble of these models can formulate a better overall prediction. Which of the following is/are true regarding weak learners used in an ensemble model?

Answers: (bold correct)

- a. They have a high bias, so they can't resolve complex machine learning problems**
- b. They usually don't overfit**
- c. They usually overfit

Question 5

True or False: The first question one should answer before getting started in machine learning development is which algorithm to use.

Answers: (bold correct)

- a. True
- b. False**

Question 6

What data type are our surface reflectance characteristics? (you can find this in the assignment for after Part 2)

Answers: (bold correct)

- a. float32
- b. int64**
- c. uint8

Question 7

Which of the following is/are basic steps in any Exploratory Data Analysis workflow? You may select more than one.

Answers: (bold correct)

- a. Data cleaning and preprocessing**

b. Statistical analysis and outlier detection

c. Visualization for trend analysis, anomaly detection, outlier detection and training

Question 8

What are the most commonly used data types for machine learning? You may select more than one.

Answers: (bold correct)

a. Numerical

b. Text

c. Boolean

Question 9

Which of the following are examples of visualization for data distribution analysis? You may select more than one.

Answers: (bold correct)

a. Histograms

b. Density plots

c. Box plots

Question 10

Which of the following are techniques for handling imbalanced data sets?

Answers: (bold correct)

a. Downsampling and bootstrapping

b. Bootstrapping and escalation

c. Downsampling and upweighting

Question 11

Which of the following is/are automated hyperparameter tuning techniques? You may select more than one.

Answers: (bold correct)

- a. Grid search
- b. Random search
- c. Bayesian search

Question 12

Which of the following is/are true for cross validation? You may select more than one.

Answers: (bold correct)

- a. **An increase in K will require more time to cross validate the result.**
- b. **Greater K values mean greater confidence in the result of cross validation compared to a lower value of K.**
- c. **If $K=N$, it's known as "Leave one out cross validation", where N is the number of observations.**

Question 13

At what point in the development of a machine learning model are interpretable machine learning explainability methods most commonly applied?

Answers: (bold correct)

- a. Before training the model
- b. While training the model
- c. **After training the model**

Question 14

True or False: Reliability is determined by interpretability and robustness.

Answers: (bold correct)

- a. **True**
- b. False

Question 15

Which of the following are disadvantages of using AutoML techniques? You may select more than one.

Answers: (bold correct)

- a. Ensemble models can provide greater generalization
- b. Ensembles consisting of several models provide less interpretability**
- c. Training and inference can be chaos computationally speaking**