Applied Remote Sensing Training (ARSET) Program

Airborne Data Applications for Invasive Species Mapping

Homework Questions

Question 1

AVIRIS-NG datasets are widely used for a variety of applications, including:

Answers: (bold correct)

a. Vegetation biodiversity
b. Greenhouse gases emissions
c. Deep bathymetric mapping
d. Both Vegetation biodiversity and Greenhouse gas emissions

Feedback:
AVIRIS-NG is a hyperspectral optical instrument mostly designed for land applications although it can also be used to study the concentration of certain atmospheric gases including methane. As an optical sensor, it has limited penetration in the water column and therefore cannot be used to construct bathymetric maps especially in deep waters.

Question 2

Which of the following Python libraries is used to access NASA data from the Earthdata cloud?

Answers: (bold correct)

a. Pandas
b. Earthaccess
c. xgboost
d. Shap

Feedback
The earthaccess library is used to access data from Earthdata cloud. Pandas, xgboost, and shap are all mentioned in the tutorial. Pandas is used for creating data frame objects. XGBoost is the name of the model and library used for the machine learning component. Shap is used to assess the performance of the XGBoost model.

Question 3

You're tasked with monitoring invasive Acacia species in a 500 km² nature reserve in South Africa. The reserve has mixed vegetation, seasonal cloud cover, and limited field access. Based on what you've learned about AVIRIS-NG capabilities and limitations, what would be your primary concerns for obtaining quality data?

A) Cloud cover
B) Mean average temperature
C) Total annual precipitation
D) time of day (day/night)
E) Cost

Answers: (bold correct)

a. E only
b. A, D, & E only
c. B & C only
d. A and C only

Feedback:
Cloud cover, time of day, and cost are all major concerns for obtaining high quality airborne hyperspectral data. Cloud cover and time of day can interfere with your ability to obtain any data at all given that the sensors record reflected solar radiation. Airborne campaigns are also very costly and it is often cost prohibitive to perform flight overpasses at a time-dense interval. The mean average temperature and total annual precipitation are not relevant to your ability to obtain quality data as these operate on much longer timescales than what is captured in a flight campaign.

Question 4

What are known benefits of utilizing machine learning in Earth Science?

A. Increased Efficiency: Machine learning can help automate the analysis of large and complex datasets, allowing scientists to quickly process and analyze large amounts of data.
B. New Insights and Discoveries: Machine learning can help scientists identify new patterns and relationships in complex datasets, leading to new insights and discoveries in Earth Science research.
C. Improved Predictive Modeling: Machine learning algorithms can be used to build accurate predictive models that can help scientists better understand complex Earth Science phenomena.

Answers: (bold correct)

a. A and B only
b. A only
c. A, B and C
d. B and C only

Feedback:
Increased efficiency, new insights/discoveries, and improved predictive models are all benefits to utilizing machine learning in Earth Science.

Question 5

As you develop a workflow for you own applications and case studies, some important considerations to keep in mind are:

A) Cost (data collection, computational resources)
B) Scale mismatch
C) Temporal resolution
D) Spectral signatures
E) Availability
F) Complexity of your project

Answers: (bold correct)

a. A, B and C only
b. A, C and D only
c. C, D and E only
d. A, B, C, D, E and F

Feedback:
The considerations listed are all important factors to keep in mind when developing your own applications of Airborne Data, and research projects in general.

Question 6

When tuning an XGBoost model for a specific application—like predicting land types over an AVIRIS scene using plot data—the hyperparameters (max_depth, learning_rate, subsample, and n_estimators) aren't set arbitrarily. Instead, they're typically selected via a tuning process that combines domain expertise with systematic experimental validation. Which hyperparameter tuning method was demonstrated in the notebook tutorial?

Answers: (bold correct)

a. Bayesian Optimization
b. Grid Search
c. Randomized Search
d. Manual Tuning

Feedback:
The hyperparameter tuning method demonstrated in the notebook was grid search (via GridSearchCV) but there are other methods for hyperparameter tuning. ALL of the possible answers listed above are valid methods for hyperparameter tuning and may be worth exploring in your own implementations of XGBoost. In Python you can use the library hyperopt for Bayesian optimization and the RandomizedSearchCV module of sklearn for randomised search. Manual tuning relies on your expertise and model performance metrics to select suitable hyperparameter settings.