Questions

Answer the following questions and discuss your answers with your classmates.

Q1) We have applied a classification technique to a hypothetical dataset with 3 classes and obtained the following results:

Actual (True) Labels	1	1	1	2	2	2	2	3	3	3
Predicted Labels	1	2	1	1	2	2	2	3	3	1

Represent the performance of this classification model with a confusion matrix and calculate the relevant evaluation metrics.

Q2) We have a dataset that contains the daily average ground-level fine particle pollution (i.e., PM2.5) measurements collected from 1479 fixed air quality monitoring stations across China during 2014–2016. In this dataset each row denotes the reading from one station at a specific date. Here are its first few rows:

	station_id	station_name	latitude	longtitude	date	PM2.5
1	001001	haidianbeibuxinqu	40.09068	116.173553	1/1/2014	11.3
2	001001	haidianbeibuxinqu	40.09068	116.173553	1/2/2014	12.1
3	001001	haidianbeibuxinqu	40.09068	116.173553	1/3/2014	12.8

Let's imagine, we have also access to spatial datasets that provide several environmental, meteorological, and land-use variables at a fine spatial resolution across China (i.e., 1 km × 1 km).

Given these datasets, can we use a machine learning model to predict daily PM2.5 at 1 km × 1 km grid cells in the entire China? Explain which type of algorithms (i.e. classification or regression) we should use, and what the training data and input and output variables are.

Python

A) Check out the scikit learn function to split matrices into random train and test subsets. Using this tool, write a notebook that reads the Iris dataset split it to train and test subsets and print the size of each subset.
B) Check out the scikit learn function to compute confusion matrix and classification metrics. Using this tool, write a notebook that takes the true and predicted labels in the first question (Q1) and calculates the confusion matrix and the classification accuracies.

Reading Material

Read the following article "Machine Learning in Agriculture: A Comprehensive Updated Review":
https://www.mdpi.com/1424-8220/21/11/3758