

GEOG3130 Advanced GIS

Assignment 4 Evaluate geospatial data quality using Python functions

Total points: 20 points

Question 1:

datalayer.csv and *truedata.csv* are the collected data and true data in a region. Perform confusion matrix calculation to evaluate the data quality (Type and value Wheat:1; Corn:2; Soy:3; Alfalfa:4; Grass:5; Fallow:6)

Suggested steps (you may develop your own steps):

1. Read both files and store data accordingly (e.g., use two lists)
 - a. You should read two files and store content with two (list) variables separately.
 - b. Each row in the file is one element in the list.
 - c. For each row, use split function to get every cell value.
 - d. Don't skip the first line.
2. After reading the content of files, compare elements in two lists and calculate correct number of cells in each category
 - a. Use a loop to retrieve every element of the list. For each loop, you should get two rows from two lists for comparisons.
 - b. Consider to use **len()** function to get the number of elements in a list.
 - c. For each row, use index to retrieve every cell.
 - d. Given the same cell location, if the cell values from two files are different, they are considered as error cells.
 - e. Make sure you convert the string values to numbers
3. Calculate the overall accuracy.
 - a. Create a variable to store the number of cells that are not correct. Whenever an error cell is detected, increase the value of the variable by 1.
 - b. Create a variable to store the number of total cells.
 - c. Those two variables should be included in the loop of Step 2.
 - d. Calculate the ratio between the number of correct cells (total-wrong) and the total number of cells.
4. Report the results in a .txt file. When you write the result, numerical values should be converted into string.

Deliverable:

1. The python script including descriptions on your code design.
2. The result produced by the evaluation script (can be a separate txt file).

Question 2:

The table *points.csv* in the assignment folder shows the control points and field observations (units: meters). According to the ASPRS standard, the RMSE accuracy should be less than **5** meters given a **1: 20,000** map, can you use the field observations as the data source to create a 1: 20,000 map?

Suggested steps (you may develop your own steps):

1. Read the data file and store the coordinate information of points.
 - a. Make sure you skip the first line which is the column header.
 - b. Create a loop to get every row in the file which is a string.
 - c. Split the string to create a list.
 - d. Use index to get values corresponding to the coordinate values from the list.
 - e. Since the values are string values, convert the strings to float.
2. Implement a function to evaluate data quality.
 - a. The function should calculate the error distance between each pair of points of a row.
 - b. The function should calculate the RMSE and return RMSE as the result.
3. Generate a report showing the results of evaluation.
 - a. You need to develop codes to determine whether the accuracy is sufficient based on RMSE value.
 - b. The report should include the value of RMSE.
 - c. The report should include the conclusion: is the data quality good enough for the map?

Deliverable:

1. The python script including descriptions on your code design.
2. The result produced by the evaluation script (can be a separate txt file).

Undergraduate students are expected to finish one question only. Graduate students are expected to finish one question and the first step of the other question (reading information---coordinates, or cells from the files).