A25: Classification with caret (2)

Your results should be delivered as an HTML webpage generated using R Markdown or Quarto. Make sure to include all code and results along with the answers to the discussion topics.

Grading Criteria

- Correctness and completeness of code (10 Points)
- Reporting of assessment metrics (10 Points)
- Correctness of interpretation of assessment metrics (10 Points)
- Quality and correctness of answers to discussion questions (10 Points)

We have provided several tables that were generated from the Eurosat dataset, which is available at the following links:

https://github.com/phelber/EuroSAT or https://www.kaggle.com/datasets/apollo2506/eurosat-dataset

These data were first introduced in this study:

Helber, P., Bischke, B., Dengel, A. and Borth, D., 2019. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(7), pp.2217-2226.

These data are designed to be used as input to CNN-based scene classification models. For this exercise, we have modified the data for use in a tabulated form. Mean spectral reflectance of all pixels included in the 64x64 pixel image extents for each sample were calculated to convert the data from image chips to a tabulated form. There are two versions of these data: EuroSat and EuroSatAllBands. We will make use of the data with all 13 bands: EuroSatAllBands. The band designations are as follows:

- 1 = Coastal Aerosol (60 m)
- 2 = Blue (10 m)
- 3 = Green (10 m)
- 4 = Red (10 m)
- 5 = Red Edge 1 (20 m)
- 6 = Red Edge 2 (20 m)
- 7 = Red Edge 3 (20 m)
- 8 = Near Infrared (NIR) (10 m)
- 8a = NIR Narrow (10 m)
- 9 = Water Vapor (60 m)
- 10 = Cirrus Cloud (60 m)
- 11 = Shortwave Infrared (SWIR) 1 (20 m)
- 12 = SWIR 2 (20 m)

These data were derived from the Multispectral Instrument (MSI) onboard the Sentinel-2A and 2B satellites, which are operated by the European Space Agency (ESA) as part of the Copernicus Program. The bands vary by spatial resolution from 10 m to 60 m; however, all bands have been resampled to a 10 m spatial resolution in the EuroSat and EuroSatAllBands datasets. We will only use the bands that were originally collected at 10 m or 20 m in this assignment: red, green, blue, red edge 1, red edge 2, red edge 3, NIR, narrow NIR, SWIR 1, and SWIR 2. A total of 10 land cover types are differentiated: annual crop, forest, herbaceous vegetation, highway, industrial, pasture, permanent crop, residential, river, and sea/lake.

For this assignment, use the train_aggregated.csv data to train models and the test_aggregated.csv data to assess the resulting models. You will not use the val_aggregated.csv data.

Task 1: Train models using the RF, SVM, and *k*NN algorithms and all provided bands. The "class" column represents the land cover classes while the band means are stored in the "blue", "green", "red", "red_edge1",

“red_edge2”, “red_edge3”, “NIR”, “NIR_Narrow”, “swir1”, and “swir2” columns. Using 5-fold cross validation, optimize relative to Kappa, test 10 values for the hyperparameters using the tuneLength argument, and center and scale the data. Use the results to predict the testing data and obtain confusion matrices and additional assessment metrics using the yardstick package. Using this package, calculate overall accuracy, class-aggregated macro-averaged precision, class-aggregated macro-averaged recall, and class-aggregated macro-averaged F1-score. Calculate class-level user’s and producer’s accuracies using the rfUtilities package.

Task 2: Train models using the RF, SVM, and *k*NN algorithms using only the red, green, and blue visible band data. Using 5-fold cross validation, optimize relative to Kappa, test 10 values for the hyperparameters using the tuneLength argument, and center and scale the data. Use the results to predict the testing data and obtain confusion matrices and additional assessment metrics using the yardstick package. Using this package, calculate overall accuracy, class-aggregated macro-averaged precision, class-aggregated macro-averaged recall, and class-aggregated macro-averaged F1-score. Calculate class-level user’s and producer’s accuracies using the rfUtilities package.

Discussion 1: Use your results to compare the different algorithms. Based on the assessment metrics, which algorithm provided the strongest performance for this task using all bands and just the red, green, and blue bands? Was the choice of algorithm more or less important when using all bands vs. just the visible bands. Explain your reasoning.

Discussion 2: Use your results and the calculated assessment metrics to compare the different feature spaces: all bands vs. just the visible bands. Did adding the additional bands improve the results? Were similar differences seen for each algorithm? Explain your reasoning.

Discussion 3: Generally, what classes were most difficult to separate or predict. Discuss this using the confusion matrices and assessment metrics.

Discussion 4: What classes saw the largest improvement when all bands were used as opposed to just the visible bands? Use the calculated assessment metrics to support your argument and conclusions.