Report for the Course Modelling in Computational Science, HT23

Project 3: Biome classification

Theo Koppenhöfer

(with Anna and Carmen, Group 4)

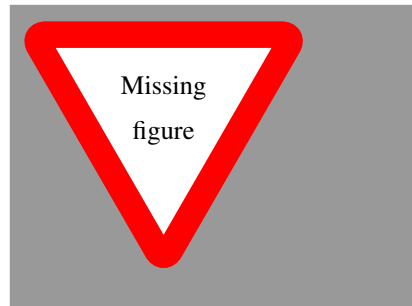Lund

October 26, 2023

Figure 1: Data points with 'desert' and 'arid shrub' in selected countries.

Introduction

The following report is part of the second project of the course Modelling in Computational Science, BERN01, taken at Lund university. In this project we will use machine learning to classify biomes based on climate and soil data. We will test the performance of our machine learning model in binary classification and in distinguishing multiple biomes for different regions. We will also compare our model with LPG_guess output and modify our model to predict continuous variables of LPG_guess. For this we will discuss the choice of regions and biomes, the setup of our model, give some interesting results, discuss these and finally give a conclusion. The code to the project was implemented in `python`. The project report and code can be found online under [1].

Methods

To test our first binary classification model we chose the biomes 'arid shrub' and 'desert'. For the choice of regions we had to choose two countries which contained sufficient amount of both regions. A plot of regions with sufficient amounts of both biomes can be seen in figure Our initial choice was Egypt and China. It turned out however that when we took out soil data our model could not handle the classification well since the deserts in both countries have very different climates. Thus we decided for Egypt for the training and Libya for the testing.

For the classification of multiple biomes we initially chose Africa and China but this quickly turned out to be a poor choice as both regions have very different climate data. Thus we switched to the regions to Russia for training and Canada for testing.

For the regression model chose Canada to train and Russia to test the model. The reason for this switch of roles lies in the performance of the training.

If not otherwise stated we use as training parameters all the parameters of the file `data_index_2.csv` excluding

```
['MaxBiomeLAI','Biome_obs','Biome_LAI','Biome_Cmax',
```

```
'Lon','Lat','Pan_2007','ISO3','UN','
                                MaxBiomeCmax
                                ',
'NPP','VegC','SoilC','LitterC','SoilR']
```

Results

Binary classification

Discussion

Conclusion

Bibliography