Exercise 3 – Machine Learning Methods

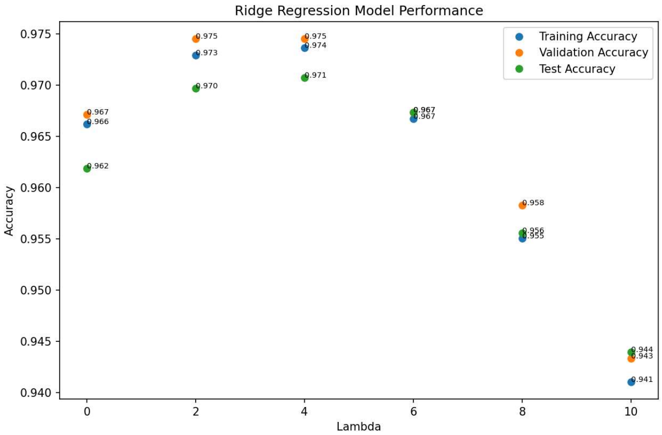
Dana Aviran

6 – Ridge Regression Task

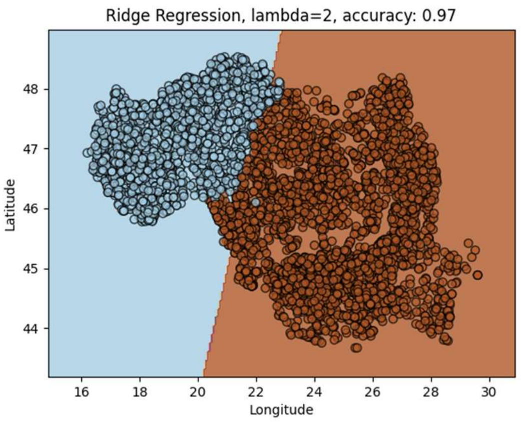
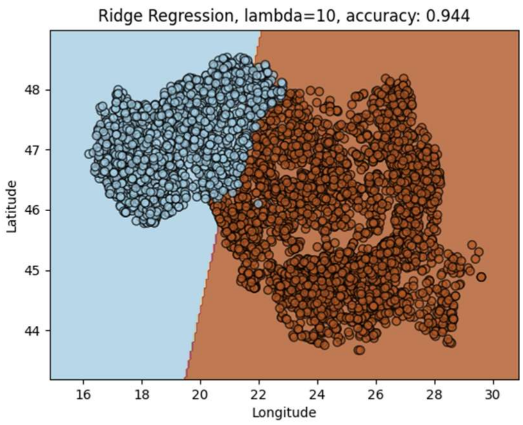
Implement a ridge regression classifier according to the skeleton file. Train that classifier on the train data with the following choices of λ: 0, 2, 4, 6, 8, 10.

Answer 1:

According to the validation set accuracies λ=2 is the best model with validation accuracy of 0.975. Thus, test accuracy of the best model according to the validation set is 0.97.



Answer 2:

<table>
<tr><th>Prediction space of best λ according to validation set (λ=2)</th><th>Prediction space of worst λ according to validation set (λ=10)</th></tr>
<tr><td>
</td><td>
</td></tr>
</table>

How does the λ parameter affect the algorithm? Explain.

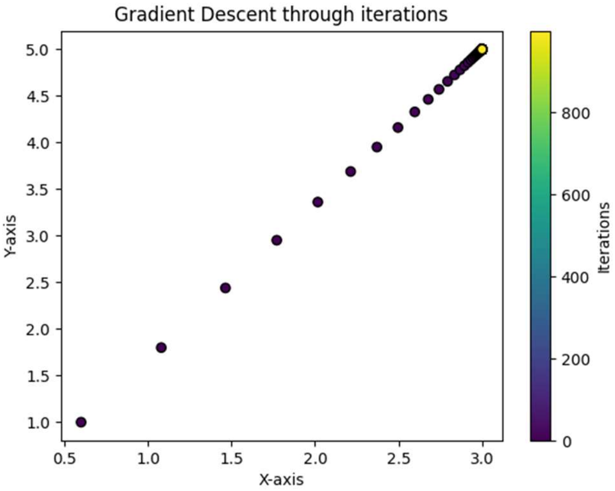
In ridge regression, the parameter lambda controls the amount of regularization applied to the model, where regularization is a technique used to prevent overfitting by penalizing large weight to eliminate unnecessary features. Our rigid regression model is trained by finding weights W that minimize the cost function: $\underset{W}{\operatorname{argmin}}\left[||XW-Y||_2^2+\lambda||W||_2^2\right]$

The lambda parameter attempts to control the trade-off between fit and complexity of the model, where the fit is measured by the first term, the Euclidean distance between the predictions and the ground truth, and the complexity is measured by the second term which is the L2 norm of the weights. Very large lambda values increase the amount of regularization applied to the model by focusing the minimization of the loss function on the second term more than on the first term, which may lead to a worse accuracy of the model and thus to underfitting. In contrast, smaller lambda values reduce regularization, allowing the model to fit the training data more closely but potentially increasing the risk of overfitting. Therefore, the lambda parameter should be not too small and not to large, as we have seen in the example above, to achieve a model with high accuracy but also as simple as possible.

7 - Gradient Descent in NumPy

Gradient descent (learning rate: 0.1) with regards to the function: $f(x, y) = (x-3)^2 + (y-5)^2$

Final Point: (2.999999999999999, 4.999999999999998)

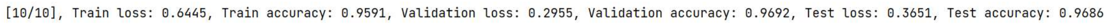


9 - Logistic Regression Classifier

Binary Case

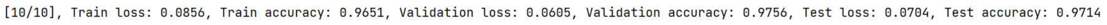
Train logistic regression classifiers on the train data with the following choices of learning rates: learning rate: 0.1, 0.01, 0.001. Use a batch size of 32 samples. Perform the training optimization for 10 epochs. Final losses and accuracies for each model (after the 10th epoch):

Learning rate of 0.1:



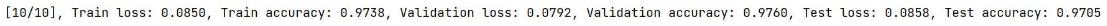
```
[10/10], Train loss: 0.6445, Train accuracy: 0.9591, Validation loss: 0.2955, Validation accuracy: 0.9692, Test loss: 0.3651, Test accuracy: 0.9686
```

Learning rate of 0.01:



```
[10/10], Train loss: 0.0856, Train accuracy: 0.9651, Validation loss: 0.0605, Validation accuracy: 0.9756, Test loss: 0.0704, Test accuracy: 0.9714
```

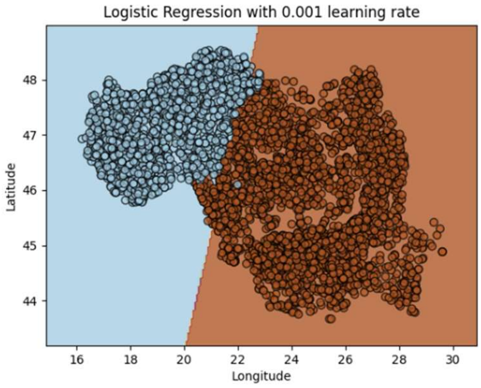
Learning rate of 0.001:



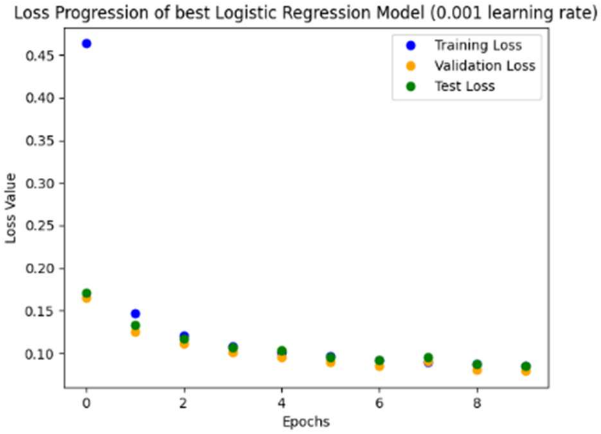
```
[10/10], Train loss: 0.0850, Train accuracy: 0.9738, Validation loss: 0.0792, Validation accuracy: 0.9760, Test loss: 0.0858, Test accuracy: 0.9705
```

Answer 1:

- Best validation accuracy: 0.976.
- Model with best validation accuracy: learning rate of 0.001.
- Test accuracy of model with best validation accuracy: 0.9705.
- Visualization of its test predictions:



Answer 2:



The graph above shows the training, validation and test loss values of the progression of the logistic regression model of learning rate 0.001 through 10 epochs, using mini-batch gradient descent. After the 10th epoch the train loss is 0.085, the validation loss is 0.0792 and the test loss is 0.0858. Although the validation loss is a little bit lower than the train and test losses, it seems that model did generalize well from the training data because the train loss is very close to the test loss and even a bit higher.

Answer 3:

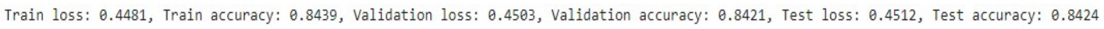
In question 1 of section 6.2, the best ridge regression model according to validation accuracy is the model where the regularization parameter equals to 2. This model has validation accuracy of 0.975 and test accuracy of 0.97. In this section, the best logistic regression classifier model according to validation accuracy is the model where the learning rate is 0.001. It has validation accuracy of 0.976 and test accuracy of 0.9705. So, it seems that both methods work well. A possible explanation is that first, the classification problem is binary, so it fits the ridge-regression classification. In addition, as we can see in the decision boundaries graph above, the decision rule for classification is linear. Thus, the ridge regression model worked well in this case although it would struggle to classify correctly in a more sophisticated classification problem that would require a decision rule that is not linear.

Multi-Class Case:

Train logistic regression classifiers on the train data with the following choices of initial learning rates: learning rate: 0.01, 0.001, 0.0003). Train your classifier for 30 epochs, with a batch size of 32. Decay the learning rate by 0.3 every 5 epochs.

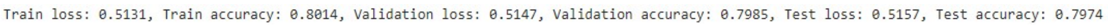
Final losses and accuracies for each model (after the 30th epoch):

Learning rate of 0.01:



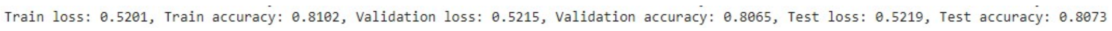
```
Train loss: 0.4481, Train accuracy: 0.8439, Validation loss: 0.4503, Validation accuracy: 0.8421, Test loss: 0.4512, Test accuracy: 0.8424
```

Learning rate of 0.001:



```
Train loss: 0.5131, Train accuracy: 0.8014, Validation loss: 0.5147, Validation accuracy: 0.7985, Test loss: 0.5157, Test accuracy: 0.7974
```

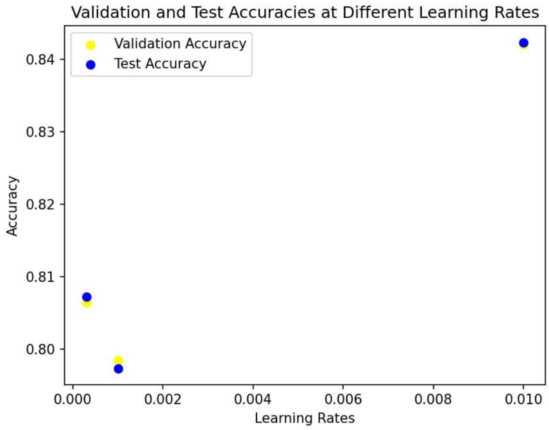
Learning rate of 0.003:



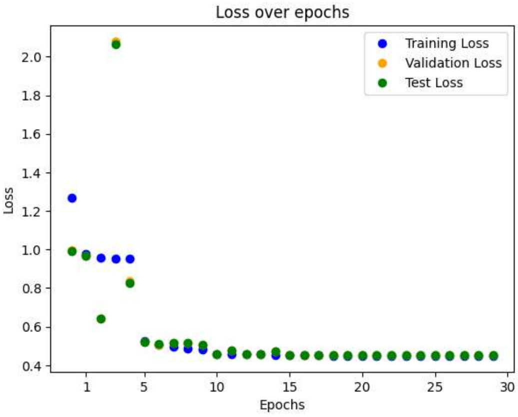
```
Train loss: 0.5201, Train accuracy: 0.8102, Validation loss: 0.5215, Validation accuracy: 0.8065, Test loss: 0.5219, Test accuracy: 0.8073
```

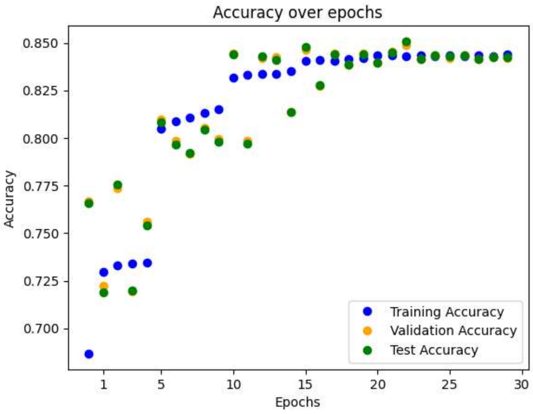
Answer 1:

According to the validation accuracy, the best model was the model of 0.01 learning rate, which has test accuracy of 0.8424. test and validation accuracies vs. their learning rate value (0.01, 0.001, 0.0003):



Answer 2:

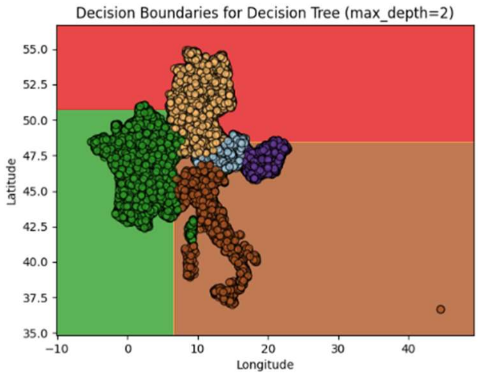




The graphs above show the training, validation and test loss and accuracy values of the progression of the logistic regression multi-class model with highest validation accuracy of learning rate 0.01 through 30 epochs, using mini-batch gradient descent. After the 30th epoch, this model has train accuracy of 0.8439, validation accuracy of 0.8421 and test accuracy of 0.8424. Moreover, we can see from the loss values graph that the loss is stable through the 15th-30th epochs, where both training, validation and test losses are very close to each-other. After the 30th epoch, the model has train loss of 0.4481, validation loss of 0.45 and test loss of 0.451. Thus, the model did generalize well from the training data because we can see that the both the validation and test accuracies and the validation and test losses are very close to the training accuracy and loss. It is worth noting that the loss seems high and that the accuracy of the multi-class logistic model after the 30th epoch is worse than the accuracy of the binary-class logistic model after the 10th epoch, and the loss is noticeably higher as well. A possible explanation might be that classification to more classes makes the problem harder to solve given the number of examples we have in our dataset or that we should have more epochs, different learning rates or different decay values.

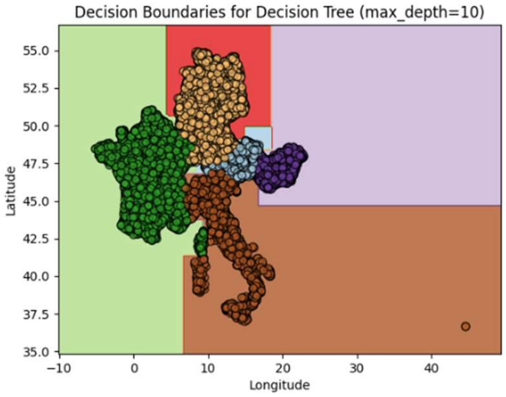
Answer 3:

Decision Tree Accuracy of max depth 2: about 0.75. The logistic regression classifier model seems better for the task since its accuracy is higher, but that is because the decision tree's max depth is 2, so the tree's expressivity is limited. A deeper tree would be more expressive and thus might be more suitable for the task since Europe countries are separated on map by axis-aligned boundaries that were originally made by human decisions, and that makes this problem suitable for decision tree classification, as we have seen in previous exercise.



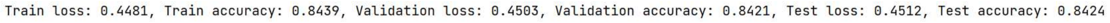
Answer 4:

Decision Tree Accuracy of max depth 10: about 0.9968. The decision tree classifier model suits better for the task – tree model of max depth of 10 is much more expressive than a tree of less depth. Moreover, as explained above, Europe countries are separated on map by axis-aligned boundaries and that makes this problem very suitable for decision tree classification.



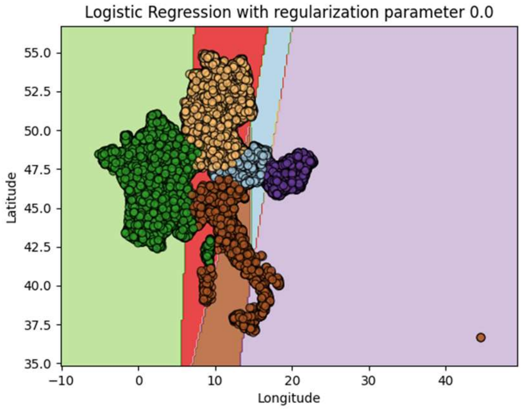
Answer 5: Bonus

Repeat the logistic regression for the logistic regression model with learning rate of 0.01 but add a ridge regularization. Use the λ values 0, 2, 4, 6, 8, 10. Final losses and accuracies for the best model where Lambda=0 (after the 30th epoch):



```
Train loss: 0.4481, Train accuracy: 0.8439, Validation loss: 0.4503, Validation accuracy: 0.8421, Test loss: 0.4512, Test accuracy: 0.8424
```

According to the validation accuracy, the best model was the model of regularization parameter=0, which has test accuracy of 0.8424. This model is the same model as the one before, which has no regularization parameter. It seems to be that the model did not need any regularization.


Logistic Regression with regularization parameter 0.0
Latitude
Longitude