Statistical Machine Learning – Week 7

Sascha Jecklin

HS 2022/23

1   Task

In this lab we will take a closer look at the implementation of Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA).

- First take a look at `week7_exercise.py`. Analyse the data using the given plot function and transform the predictiors into SI units.

- Use the `sklearn.discriminant_analysis` module to generate and fit a LDA/QDA model to the training data.

- Calculate the misclasification rate on training- and test data using the LDA and the QDA model.

- Study the output by using the given plot function. What do you notice? Run the script several times and observe the changes.

- In the `my_lda.py` we make our own implementation of a LDA model. For that study equation 4.19 in the textbook.

- Bonus exercise: Implement your own `QDA` class.

Comments

- To calculate the co-variance matrix take a look at numpy's `cov()` function `https://docs.scipy.org/doc/numpy/reference/generated/numpy.cov.html` as well as equation 4.15 on page 141 in the textbook.

- Make sure that you weight each class covariance matrix according to the number of observations in the corresponding class.

$$\Sigma = \frac{(n_1 - 1)\Sigma_1 + (n_2 - 1)\Sigma_2 + ... + (n_k - 1)\Sigma_k}{(n_1 - 1) + (n_2 - 1) + ... + (n_k - 1)}$$

- A pound is 0.45359237 kg.

- An inch is 2.54 cm.