DS6040: Homework 2 (module 3.5): Bayes Classifiers as Linear Discriminants

Posterior Probabilities with LDA and QDA

Diana McSpadden (hdm5s)

Instructions: You may discuss this assignment with other students in the class, but you must submit your own answers to the questions below. Include an honor pledge with your submission. Submit on-line and in pdf. This exercise is ungraded.

Q1

Download the required files:

- Linear Discriminant Analysis.py,
- Quadratic Discriminant Analysis.py, and
- iris data.csv

into a single folder, and watch the video explaining the Python code.

Q2

The code for the LDA and QDA classes are the same except for two lines:

1. one line in the init method and,
2. one in the compute likelihoods method

What are these two lines, and what is the functional/mathematical differences in what they do.

Q2 Answer:

init method difference:

The QDA file calculates the covariance for each of the classes for the data; however,

The LDA file sets the covariance for all classes to the mean covariance:

self.cov = np.zeros([self.num_cols, self.num_cols]), and then self.cov is set to a single value for all the class names by making it the mean covariance of the classes to the data.

This makes the assumption that the variance-covariance for all classes is equal, and only the mean **values** change, not their variance.

compute likelihoods method difference:

The QDA file uses the array of covariances for each class to calculate the PDF, while the LDA file uses the mean covariance for all classes.

Q3

In each of the Python files, complete the code for the method compute probabilities, and use your resulting code to answer the following two questions. (Submit your code with your solution to this exercise.)

In [5]:
```python
# code investigating the iris data
import numpy as np
import pandas as pd
df = pd.read_csv(r'C:\Users\dianam\Documents\Personal\BayesianML\Mod3\iris_data.csv')
df.head()#
df.groupby('Class').mean()
```

Out[5]:

	Sepal Length	Sepal Width	Petal Length	Petal Width
Class				
Iris-setosa	5.006	3.418	1.464	0.244
Iris-versicolor	5.936	2.770	4.260	1.326
Iris-virginica	6.588	2.974	5.552	2.026

- The first sample in the default LDA and QDA code is [5.1, 3.5, 1.4, 0.2], which almost looks like a prototypical iris-setosa.
- Q4a is [5.5, 2.4, 3.8, 1.1] which I am guessing will be an iris-versicolor.
- Q4b is [5.5, 3.1, 5, 1.5], I'm not sure how this one will be predicted.

Q4

Your friend Carl calls you and says he has measured two iris flowers and wants your help determining the species.

The first flower has dimensions:

- [SepalLength, SepalWidth,PetalLength,PetalWidth] = [5.5, 2.4, 3.8, 1.1]

and the second has dimensions:

- [5.5, 3.1, 5, 1.5].

Assuming noninformative priors, what are the most probable species according to LDA and QDA, and what are the probabilities for each species.

Answers

1. [5.5, 2.4, 3.8, 1.1]
   - QDA: **Iris-versicolor**
     - Iris-setosa: 2.711263447644606e-52
     - **Iris-versicolor: 0.9999701265523081**
     - Iris-virginica: 2.987344769195638e-05
   - LDA: **Iris-versicolor**
     - Iris-setosa: 1.0221808316494009e-17
     - **Iris-versicolor: 0.9999970568617268**
     - Iris-virginica: 2.9431382732285363e-06
2. [5.5, 3.1, 5, 1.5]
   - QDA: **Iris-virginica**
     - Iris-setosa: 2.726593926101304e-103
     - Iris-versicolor: 0.43517519384112463
     - **Iris-virginica: 0.5648248061588753**
   - LDA: **Iris-versicolor**
     - Iris-setosa: 2.4269307963452444e-28
     - **Iris-versicolor: 0.5602348418832765**
     - Iris-virginica: 0.43976515811672334

Q5

You realize Carl is working in the country of Bagend, and in this country 70% of the irises are Iris-virginica, 20% are Iris-versicolor, and 10% are Iris-setosa.

Use this information to create informative priors, and use your python code to provide an updated answer to the previous question.

Answers

1. [5.5, 2.4, 3.8, 1.1]
   - QDA: **Iris-versicolor**
     - Iris-setosa: 1.3555304878994901e-52
     - **Iris-versicolor: 0.9998954507411951**
     - Iris-virginica: 0.00010454925880481196
   - LDA: **Iris-versicolor**
     - Iris-setosa: 5.110866553279598e-18
     - **Iris-versicolor: 0.9999896990918361**
     - Iris-virginica: 1.030090816380721e-05
2. [5.5, 3.1, 5, 1.5]
   - QDA: **Iris-virginica**
     - Iris-setosa: 5.651997976619857e-104
     - Iris-versicolor: 0.1804162542518483
     - **Iris-virginica: 0.8195837457481517**
   - LDA: **Iris-virginica**
     - Iris-setosa: 5.780022600099141e-29

- Iris-versicolor: 0.26685310123590844
- **Iris-virginica: 0.7331468987640916**