

HW #5: Multivariate Methods

Due: 10/13/21 (Wednesday) 11:59PM. **DO NOT COPY OTHER'S WORK.**

This problem is based on an online article: [Naive Bayes Classifier From Scratch in Python \(machinelearningmastery.com\)](https://machinelearningmastery.com/naive-bayes-classifier-from-scratch-in-python/). You should read the article carefully and learn how to implement a Naïve Bayesian classifier in Python. Note that in a Naïve Bayesian classifier, it is assumed that the features are independent and thus the covariance matrix is diagonal. Your task in this assignment is to modify the given code to implement a Multivariate Bayesian classifier that uses a general covariance matrix.

1. Learn how k-fold cross-validation is implemented to evaluate the performance of the algorithm in the provided code. In the code, 5 folds are used. Run the provided code and check the resulting accuracy of the algorithm. Note that 5-fold cross-validation produces 5 different accuracies and you need to check and report the mean accuracy as well as the 5 different accuracies.
2. Copy the provided code and change the file name to ***HW5\_Iris\_Multi\_YourLastName.py***. In the new file, change the function name *naïve\_bayes* to ***multivariate\_bayes***. Also, to incorporate a general covariance matrix with nonzero off-diagonal elements, modify the internal code of the following functions: ***calculate\_class\_probabilities***, ***calculate\_probability***, and ***summarize\_dataset***. Finally, create a new function named ***cov*** to calculate the covariance matrix. Note that your specific implementation may be different from mine and thus the actual functions that you modify and create may be different from what are listed above. Copy and paste the functions that you modified and created in your report.

Note: You must implement your own algorithm for the calculation of a covariance matrix. Do not use the built-in function, `numpy.cov()` or any similar function. You are allowed to use built-in functions for the matrix addition/subtraction, multiplication, determinant, transpose, and inverse matrix calculation.

3. Run your new code and report the five accuracies and mean accuracy. Also, submit ***HW5\_Iris\_Multi\_YourLastName.py*** along with the report. Does the general Multivariate Bayesian Classifier show improved performance than the Naïve Bayesian Classifier?