Structure of Code:

- The code is designed using Python and contains a function for classification problem
- First, the input file “iris.data” is read as a dataframe using numpy library.
- Then the three iris species names are changed to ‘0’,’1’,’2’ labels in float format and the entire data is manually split into training and testing data(80/20)
- With training data as X and training labels as Y we find Beta cap β̂ using linear regression model formula,

$$\hat{\beta} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{Y}$$

- Now, using the estimator values we classify the model and also predict label values over the testing data,

$$f(\mathbf{x}) = w_0 + w_1x_1 + w_2x_2 + \ldots w_dx_d = w_0 + \sum_{j=1}^{d} w_jx_j$$

- In order to cross validate we fit the data again into predefined regression model and Compute the predicted values from N-fold cross validation and pass it in regression score function to get the accuracy of predicted output which comes out to be 1.0 mostly.
- Since there is no discrepancy in no. of folds, the N value is chosen to be **10** which is most widely used as the more folds we have, we will be reducing the error due to the bias but increasing the error due to variance which nullifies each other.

Screenshot of output: