

## Short Report on Naive Bayes Classifier Implementation

### Data Processing Steps

#### 1. Data Loading:

- The Iris dataset was loaded using `sklearn.datasets`. This dataset consists of 150 samples of iris flowers, with four features: sepal length, sepal width, petal length, and petal width. The target variable is the species of the iris (Setosa, Versicolor, or Virginica).

#### 2. Feature Selection:

- For this analysis, only two features were selected: **petal length** and **petal width**. This simplification helps visualize the decision boundaries effectively in a two-dimensional plot.

#### 3. Data Splitting:

- The dataset was split into training and testing sets using `train_test_split` from `sklearn.model_selection`. A test size of 20% was chosen to validate the model's performance.

### Model Choice

#### • Naive Bayes Classifier:

- A Gaussian Naive Bayes classifier was chosen for this task due to its simplicity and effectiveness in handling continuous data, such as the petal dimensions in the Iris dataset. The Naive Bayes algorithm assumes that features are independent given the class label, which often holds true in real-world scenarios.

### Performance Evaluation

#### 1. Model Training:

- The Naive Bayes model was trained using the training dataset with the `fit` method. This process involved calculating the mean and variance of each feature for each class.

## 2. Decision Boundaries Visualization:

- The decision boundaries were visualized using a contour plot, where the model's predictions across a grid of petal lengths and widths were represented. Each region in the plot corresponds to a predicted class label, illustrating how the model distinguishes between the three species.

## 3. Class Labels:

- A legend was added to the plot to indicate which color corresponds to which species. This aids in interpreting the decision boundaries in relation to the actual flower classes.

## Insights Gained

- The decision boundary plot reveals how well the Naive Bayes model separates the three classes based on the selected features.
  - The plot indicates:
    - **Setosa** is well-separated from both **Versicolor** and **Virginica**.
    - **Versicolor** and **Virginica** have some overlap, suggesting that the model may have difficulty distinguishing between these two classes based solely on petal dimensions.
- The simplicity of the Naive Bayes classifier, combined with the clear separation observed in the plot, showcases its potential effectiveness for this type of classification problem.