Machine Learning HW1

Shang-Ying He
Communication Engineering
sunnyhe1122@gmail.com

In this data set we have four features for each data and for each feature I denoted as f_1, f_2, f_3, f_4 respectively.

For the maximum a posterior probability, target is to compute the result of posterior probability (1) for each class and select the maximum as predict, after that we can get (2) (3) with simple derivation. And $P(f_1|C_i)P(f_2|C_i)P(f_3|C_i)P(f_4|C_i)$ called likelihood function and $\mathrm{P}(C_i)$ is prior probability.

$$\mathrm{P}(C_i|\text{features}) = \frac{P(C_i \cap features)}{P(features)} \qquad C_i\text{: class i} \tag{1}$$

$$\begin{aligned}\mathrm{P}(C_i \cap \text{features}) &= \mathrm{P}(C_i \cap \mathrm{f}_1)\mathrm{P}(C_i \cap \mathrm{f}_2)\mathrm{P}(C_i \cap \mathrm{f}_3)\mathrm{P}(C_i \cap \mathrm{f}_4) \\ &= \mathrm{P}(C_i)^4 P(f_1|C_i)P(f_2|C_i)P(f_3|C_i)P(f_4|C_i)\end{aligned} \tag{2}$$

$$\begin{aligned}\mathrm{P}(\text{features}) &= \sum_i \mathrm{P}(C_i \cap \mathrm{f}_1)\mathrm{P}(C_i \cap \mathrm{f}_2)\mathrm{P}(C_i \cap \mathrm{f}_3)\mathrm{P}(C_i \cap \mathrm{f}_4) \\ &= \mathrm{P}(\text{features} \cap C_1) + \mathrm{P}(\text{features} \cap C_2) + \mathrm{P}(\text{features} \cap C_3)\end{aligned} \tag{3}$$

So, what we need to do first is to compute prior probability for each class from training data, in this training set I got the prior as (4) for each class respectively.

$$\begin{cases} Iris - setosa = 0.325 \\ Iris - versicolor = 0.325 \\ Iris - virginica = 0.35 \end{cases} \tag{4}$$

And the other thing we need is to compute the mean and standard deviation between classes and features from training data, TABLE I. blow is the result I computed.

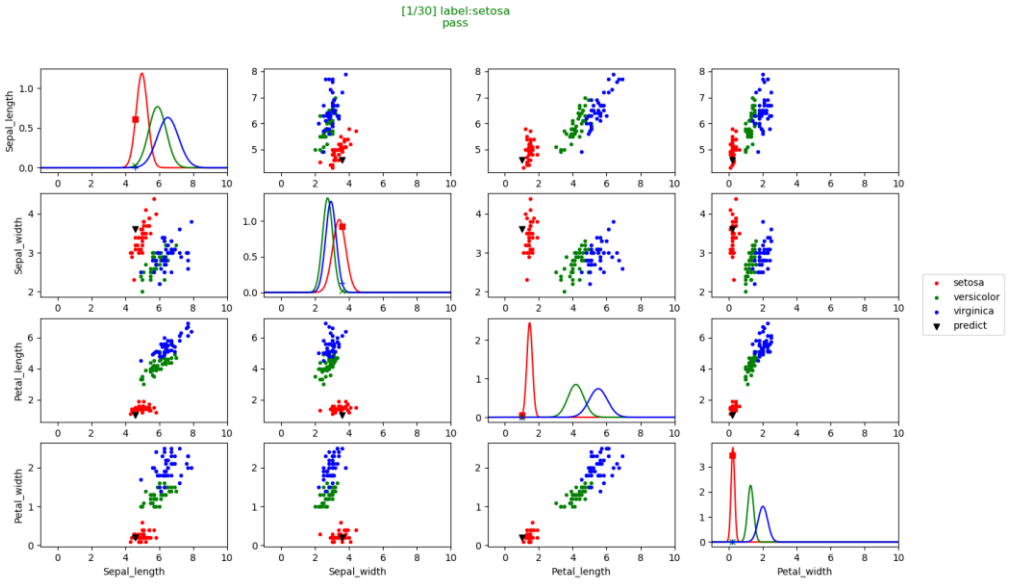
So far, we have all requirements to compute likelihood, just to follow the mean and standard deviation in TABLE I. to estimate a Gaussian distribution from each testing data.

Final step is to compute $\mathrm{P}(\text{features})$, after getting the result of likelihood it is very simple to do, just to compute sum of likelihood product with prior from every class like function (3).

TABLE I					
Class \ Feature		Sepal length	Sepal width	Petal length	Petal width
setosa	Mean	4.9871793	3.4230766	1.4564103	0.2435897
	Standard deviation	0.3352607	0.39122528	0.16297694	0.10572065
versicolor	Mean	5.907693	2.7435896	4.184615	1.2897437
	Standard deviation	0.51758045	0.30195504	0.4698818	0.17656967
virginica	Mean	6.5190473	2.9357142	5.502381	2.0142856
	Standard deviation	0.6299047	0.3130658	0.54005647	0.2807909

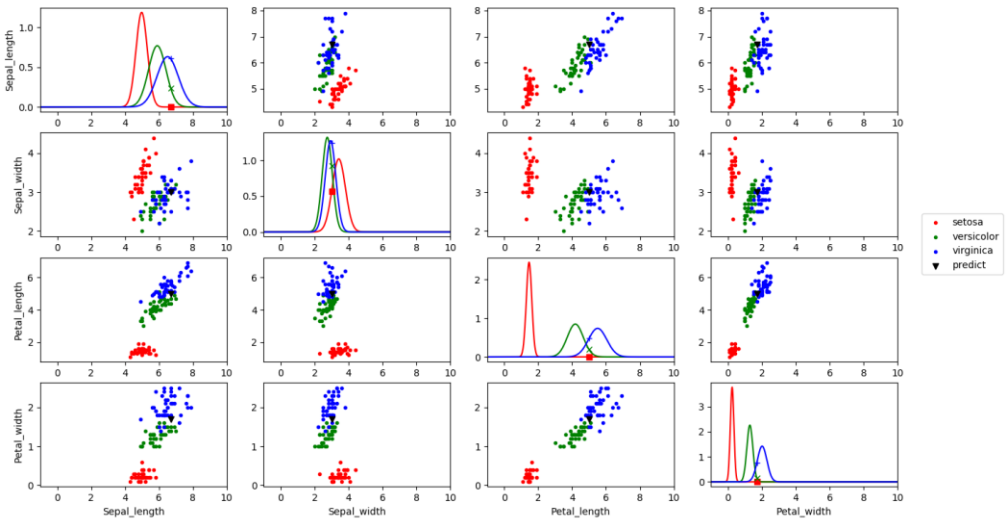
Above of all, I got the training error rate is 0.0083333 and testing error rate is 0.0666666. In other words, in 30 piece of testing data, I got two errors to predict.

Figure blow is an example of predict correctly, we can observe that for each feature testing data estimate in Gaussian is always distributing in class “setosa”.



Figures blow are two errors in testing data, all of them have the same characteristic which is distributed at the overlap of Gaussian where is always causing the error between each class, if we want to reduce the error rate in this task, we need to use another distribution to estimate the data which can make three classes’ distribution is more sparse than Gaussian.

[9/30] label:versicolor
predict:virginica
error



[20/30] label:versicolor
predict:virginica
error