

Часть 4

Задания:

4.1. Построить байесовский классификатор в предположении, что переменная распределена нормально для обоих классов (5 баллов).

4.2. Оценить точность классификации при случайном разбиении выборки на

обучающую (80%) и контрольную выборки (20%) (5 баллов).

Код:

```
from scipy import stats as sps
import pandas as pd
from sklearn.model_selection import *
from math import *
import random
random.seed(993)

data = pd.read_csv("iris.data")
X = "Iris-versicolor"      # choose your parameter here
Y = "Iris-setosa"          # choose your parameter here
TYPE = "sepal_length"      # choose your parameter here

def classes(data, X_name, Y_name, TYPE):
    # get data
    X = data[data["species"] == X_name][TYPE]
    Y = data[data["species"] == Y_name][TYPE]
    # get train/test data
    train1, test1 = train_test_split(X, test_size=0.2, random_state=3)
    train2, test2 = train_test_split(Y, test_size=0.2, random_state=3)

    # sample average for the classes X and Y
    mean1 = train1.mean()
    mean2 = train2.mean()
    # sample variance for the classes X and Y
    var1 = ((train1 - mean1) ** 2).sum() / train1.size
    var2 = ((train2 - mean2) ** 2).sum() / train2.size

    print("m1 == ", mean1, ", m2 == ", mean2)
    print("var1 == ", var1, ", var2 == ", var2)

    pred_results = {"Positive": 0, "Negative": 0}
    # the number of correct predictions for X and Y
    for i in test1:
        if sps.norm(loc=mean1, scale=sqrt(var1)).pdf(i) > sps.norm(loc=mean2,
scale=sqrt(var2)).pdf(i):
            pred_results["Positive"] += 1

    for i in test2:
        if sps.norm(loc=mean1, scale=sqrt(var1)).pdf(i) < sps.norm(loc=mean2,
scale=sqrt(var2)).pdf(i):
            pred_results["Negative"] += 1
```

```
print("Result of prediction for Positive: ", pred_results["Positive"])
print("Result of prediction for Negative: ", pred_results["Negative"])

# accuracy of the Bayes classifier
acc = (pred_results["Positive"] + pred_results["Negative"]) / (len(test1)
+ len(test2))

print("classifier accuracy =", f"{acc * 100}%")

return acc

classes(data, X, Y, TYPE)
```

Вывод:

m1 == 5.9799999999999995 , m2 == 5.0325000000000001

var1 == 0.2766 , var2 == 0.12119374999999999

Result of prediction for Positive: 8

Result of prediction for Negative: 10

classifier accuracy = 90.0%

Ответ: 0.9 или 90%