Міністерство освіти і науки України

Харківський національний університет радіоелектроніки

Кафедра Штучного інтелекту

Звіт з лабораторної роботи №5

з курсу Машинне навчання

на тему “ Реалізація та дослідження алгоритму Naïve Bayes ”

Виконав:

студент гр. ІТШІ-17-1

Гура А.О.

Прийняла:

Кулішова Н.Є.

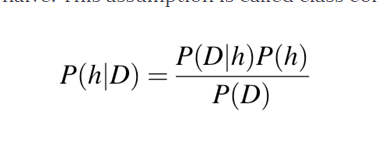
Харків 2019

**1.Цель работы:** Закріпити основні поняття імовірнісного навчання. Застосувати алгоритм Naïve Bayes для класифікації нових прикладів.

**2.Описание алгоритма:**

Наи́вный ба́йесовский классифика́тор — простой вероятностный классификатор, основанный на применении теоремы Байеса со строгими (наивными) предположениями о независимости.

В зависимости от точной природы вероятностной модели, наивные байесовские классификаторы могут обучаться очень эффективно. Во многих практических приложениях для оценки параметров для наивных байесовых моделей используют метод максимального правдоподобия; другими словами, можно работать с наивной байесовской моделью, не веря в байесовскую вероятность и не используя байесовские методы.

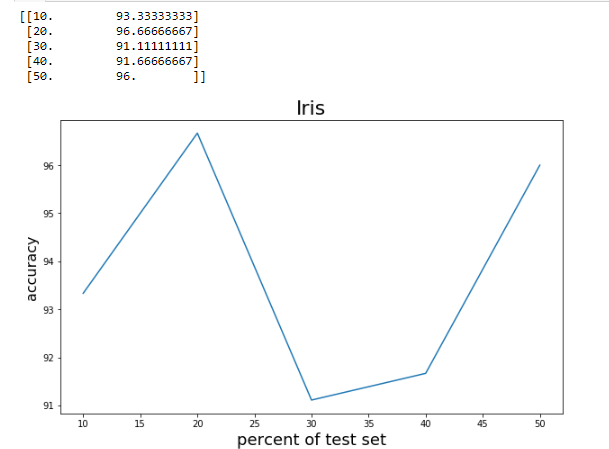
Основная формула:

**3.Ход работы:**

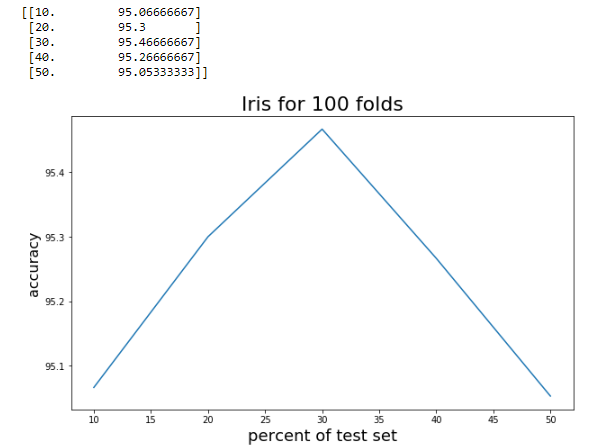
Во время выполнения лабораторной работы я использовал для испытаний алгоритма три выборки Iris, Wine и Breast Cancer. Для каждой из выборок я запускал алгоритм для различных разбиений на тестовую/тренировочную выборки (50/50, 60/40, 70/30, 80/20, 90/10). А также сохранял значения точности и строил для них графики.

Также хочу заметить что я пользовался классификатором из библиотеки sklearn GaussianNB.

**4. Графические зависимости:**

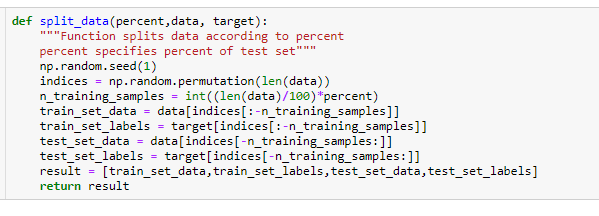
**** - График для Ириса на 1 цикл:

-График для Ириса на 100 циклов со средней точностью

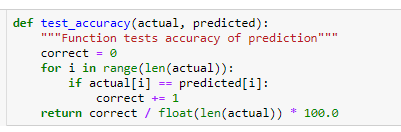


**5. Тексты скриптов:**

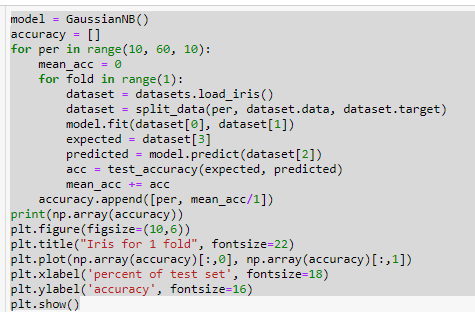
- Функция для разбиения выборки и перемешки

****

- Для оценки точности



- Для классификации Ириса



**6. Выводы**

В ходе выполнения данной лабораторной работы я ознакомился с алгоритмом Naïve Bayas а также на практике протестировал его возможности на трех различных датасетах.

Кроме того, я увидел влияние как изменяется точность классификации в зависимости от размера выборки для тестирования, а именно что если брать точность только за 1 прогон алгоритма, то она может крайне сильно изменятся (от 100 до 80), ну а если брать среднюю точность за 100 прогонов, то она практически идентичная на любом разбиении.