ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ

ВЫСШЕГО ОБРАЗОВАНИЯ

***«*САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ПЕТРА ВЕЛИКОГО»**

Институт компьютерных наук и технологий

**Высшая школа программной инженерии**

Изображение выглядит как текст

Автоматически созданное описание

**ЛАБОРАТОРНАЯ РАБОТА №4**

по дисциплине «Машинное обучение»

Студент А. М. Потапова

гр. 3530202/90202

Руководитель И. А. Селин

Санкт-Петербург

2022 г

**Содержание**

[Задание 1 3](#_Toc117671930)

[Задание 2 5](#_Toc117671931)

[Задание 3 7](#_Toc117671932)

# **Задание 1**

Исследуйте зависимость качества классификации от количества классификаторов в ансамбле для алгоритмов бэггинга на наборе данных glass.csv с различными базовыми классификаторами. Постройте графики зависимости качества классификации при различном числе классификаторов, объясните полученные результаты.

*Исходные данные:*

Изображение выглядит как текст

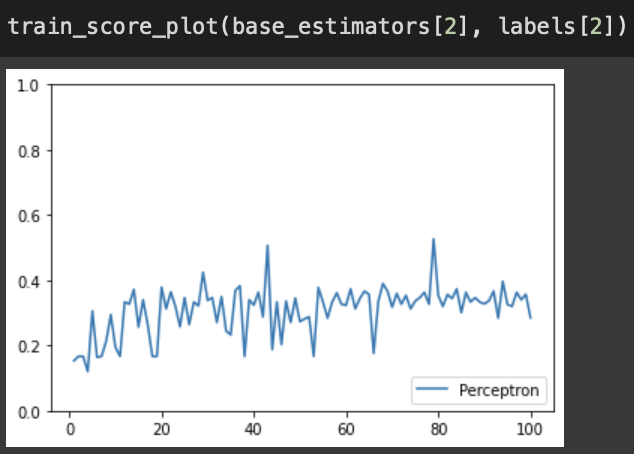
Автоматически созданное описание

Алгоритм бэггинга – BaggingClassifier

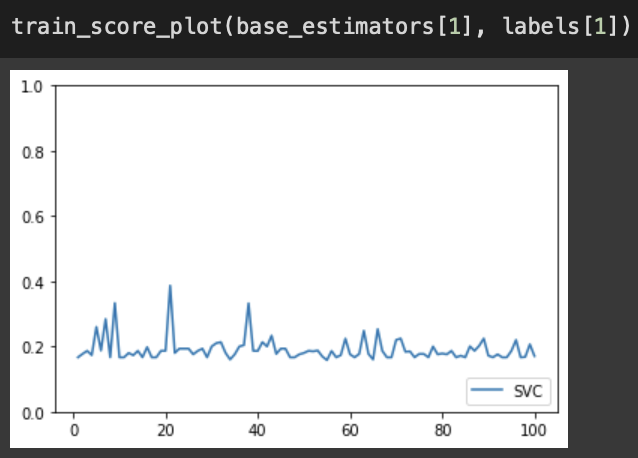
Базовые классификаторы – Perceptron, SVC и DecisionTreeClassifier

Метрика классификатора – balanced\_accuracy\_score

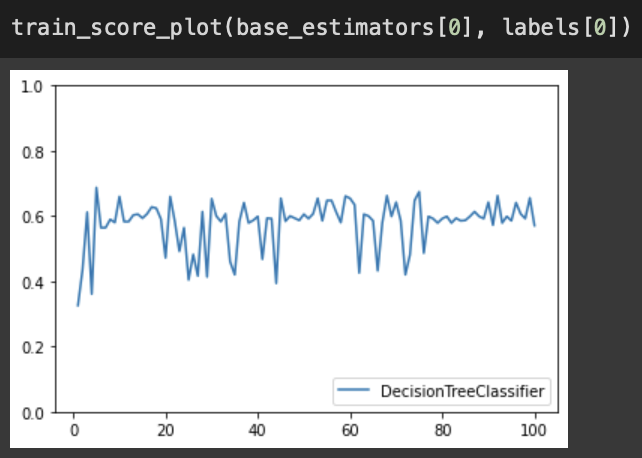
* Используем классификатор Perceptron. Зависимость значения метрики от количества классификаторов:



* Используем классификатор SVC. График зависимости:



* Используем классификатор DecisionTreeClassifier. График зависимости:



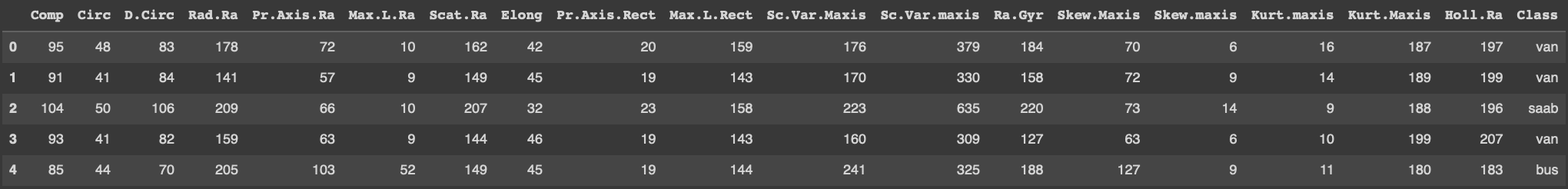
**Вывод**

Исходя из полученных результатов можно отметить, что среди этих 3 классификаторов наилучшие значения метрики balanced\_accuracy\_score показал классификатор DecisionTreeClassifier, т. к. его качество классификации в среднем равна 0.6. Для него же наблюдаем периодический рост качества классификации в виде часто встречающихся пиков. При использовании алгоритмов бэггинга с ростом числа классификаторов видим несущественный рост качества классификации для базового классификатора Perceptron. У SVC не видим существенного роста качества классификации, а лишь довольно редкие пики.

# **Задание 2**

Исследуйте зависимость качества классификации от количества классификаторов в ансамбле для алгоритма бустинга (например, AdaBoost) на наборе данных vehicle.csv с различными базовыми классификаторами. Постройте графики зависимости качества классификации при различном числе классификаторов, объясните полученные результаты.

*Исходные данные:*

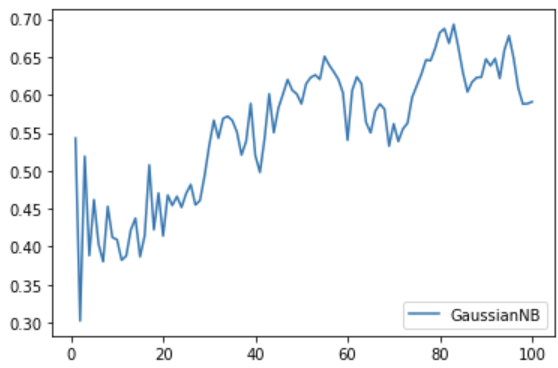


Алгоритм бустинга – AdaBoostClassifier

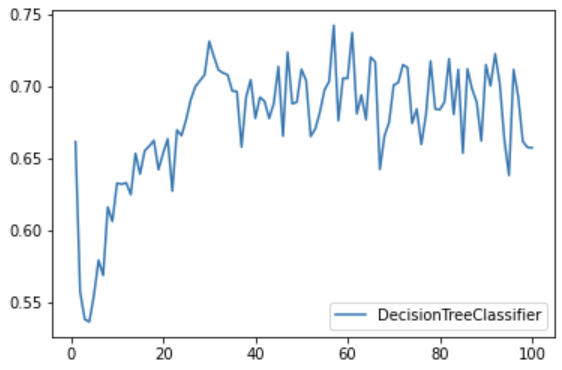
Базовые классификаторы – GaussianNB и DecisionTreeClassifier

Метрика классификатора – balanced\_accuracy\_score

* Используем классификатор GaussianNB. Зависимость значения метрики от количества классификаторов:



* Используем классификатор DecisionTreeClassifier. График зависимости:



**Вывод**

В обоих случаях видим сначала резкий спад качества классификации, а затем значительный рост. Базовый классификатор DecisionTreeClassifier оказался немного лучше базового классификатора GaussianNB, т. к. качество классификации у него выше (больше 0.7). Оба случая показали достаточно большие значения метрики balanced\_accuracy\_score.

# **Задание 3**

Постройте мета-классификатор для набора данных titanic\_train.csv используя стекинг и оцените качество классификации на titanic\_train.csv

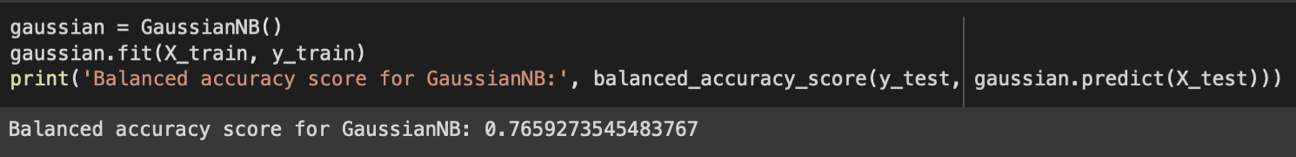
*Исходные данные:*

Изображение выглядит как текст, монитор, снимок экрана

Автоматически созданное описание

Базовые классификаторы – GaussianNB, SVC, KNeighborsClassifier и DecisionTreeClassifier

* Используем классификатор GaussianNB. Полученные значения метрики balanced\_accuracy\_score:



*Результат для GaussianNB: 0.7659273545483767*

* Используем классификатор SVC. Полученные значения метрики balanced\_accuracy\_score:

Изображение выглядит как текст

Автоматически созданное описание

*Результат для SVC: 0.7524911603985857*

* Используем классификатор KNeighborsClassifier. Полученные значения метрики balanced\_accuracy\_score:

Изображение выглядит как текст

Автоматически созданное описание

*Результат для KNeighborsClassifier: 0.7482802957248473*

* Используем классификатор DecisionTreeClassifier. Полученные значения метрики balanced\_accuracy\_score:

Изображение выглядит как текст

Автоматически созданное описание

*Результат для DecisionTreeClassifier: 0.7151076824172292*

* Теперь мы можем узнать значение balanced\_accuracy\_score для мета-классификатора:

Изображение выглядит как текст

Автоматически созданное описание

*Результат для мета-классификатора: 0.803953712632594*

**Вывод**

Исходя из полученного результата можно отметить, что качество классификации при построении мета-классификатора, используя стекинг, возросло. При использовании отдельных классификаторов среднее значение balanced\_accuracy\_score составило 0,7454, а в случае с мета-классификатором – 0,8.