

Министерство образования и науки Российской Федерации

САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ПЕТРА
ВЕЛИКОГО



ПОЛИТЕХ

Санкт-Петербургский
политехнический университет
Петра Великого

Отчет по Лабораторной работе № 4.
по дисциплине “Машинное обучение”

Выполнила
студентка гр. 3530202/00201

Руководитель

Козлова Е. А.

Селин И. А.

Санкт-Петербург
2023

Оглавление

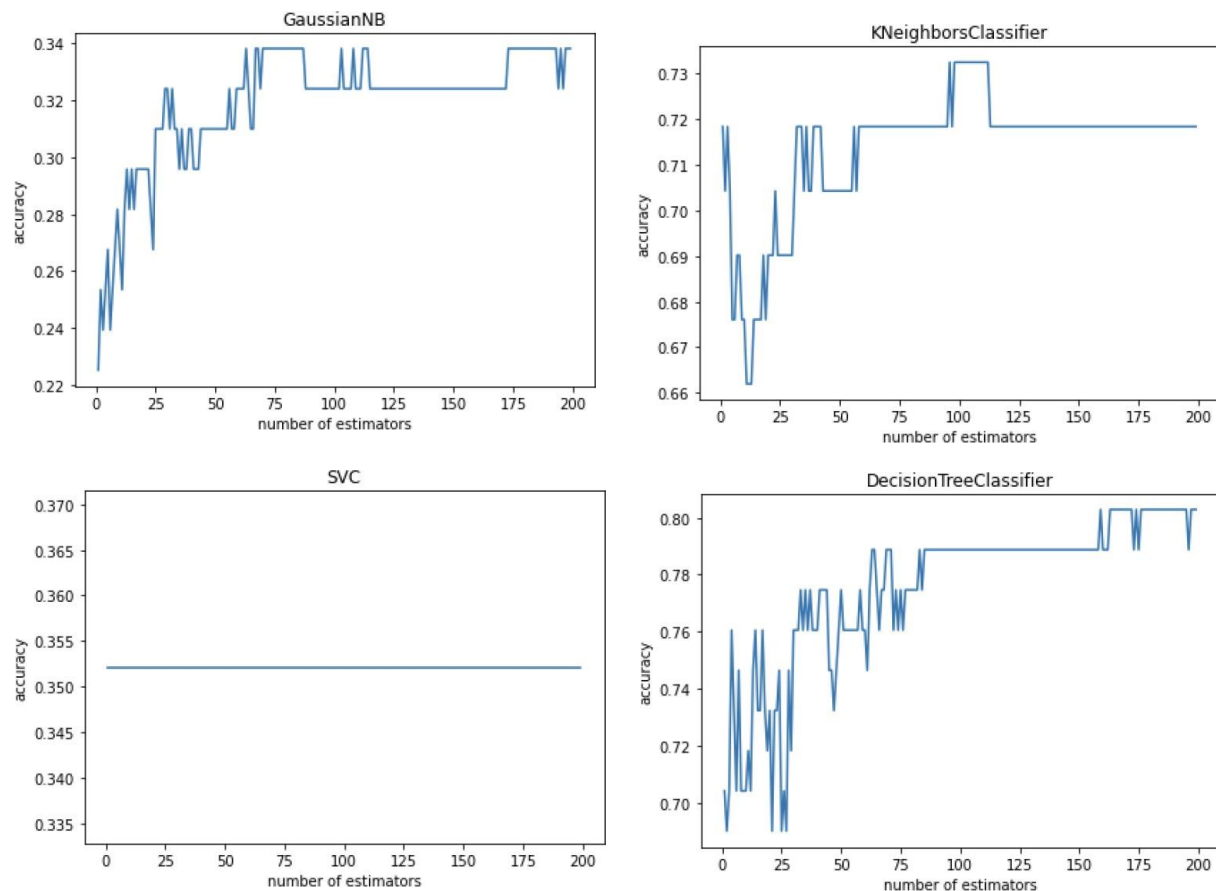
Задание 1.....	3
Задание 2.....	4
Задание 3.....	5

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.

Была проведена анализ зависимости точности классификации от числа классификаторов в ансамбле при использовании алгоритмов бэггинга на данных из файла glass.csv. Этот анализ включал различные базовые классификаторы, и результаты были визуализированы в виде графиков, демонстрирующих влияние количества классификаторов на качество классификации.

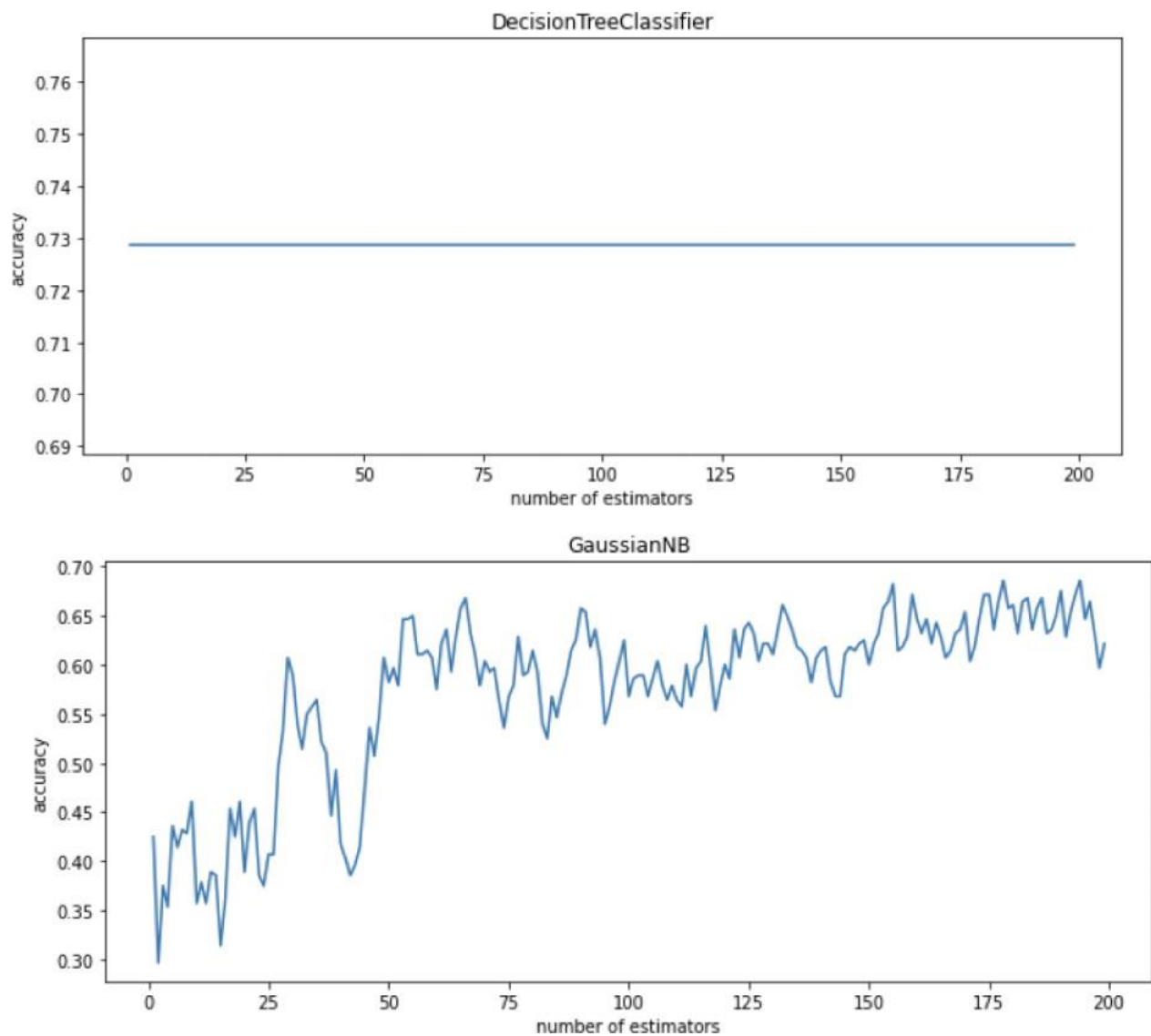
В ходе работы были рассмотрены следующие классификаторы - GaussianNB, KNeighborsClassifier, SVC, DecisionTreeClassifier:



Вывод: как мы можем видеть из полученных графиков, алгоритму бэггинга удалось добиться большей точности, чем если бы мы использовали классификаторы отдельно. Но у нас не получилось улучшить SVC.

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.



Вывод: из представленных диаграмм видно, что байесовский классификатор значительно улучшился при использовании алгоритма бустинга. Тем не менее, с точки зрения точности классификации, дерево решений превосходит байесовский классификатор, даже несмотря на то, что оно остается неизменным и не подвергается улучшениям.

Задание 3

Постройте мета-классификатор для набора данных `titanic_train.csv` используя стекинг и оцените качество классификации на `titanic_train.csv`

Точность для отдельных классификаторов:

```
"Naive Bayes" = 0.7790571846802059
"K neighbours" = 0.7990357841048976
"SVC" = 0.7790571846802059
"Decision tree" = 0.7990357841048976
```

Точность для комбинации двух классификаторов:

```
"Naive Bayes" + "K neighbours" = 0.8112504603578413
"Naive Bayes" + "SVC" = 0.800368756320145
"Naive Bayes" + "Decision tree" = 0.7988632011276058
"K neighbours" + "SVC" = 0.7963578521350468
"K neighbours" + "Decision tree" = 0.7898305084745763
"SVC" + "Decision tree" = 0.7886220145346907
```

Точность для комбинации трех классификаторов:

```
"Naive Bayes" + "K neighbours" + "SVC" = 0.8067852036721302
"Naive Bayes" + "K neighbours" + "Decision tree" = 0.7963491078393045
"Naive Bayes" + "SVC" + "Decision tree" = 0.7863501285326791
```

Точность для комбинации четырех классификаторов:

```
"Naive Bayes" + "K neighbours" + "SVC" + "Decision tree" = 0.8033898305084746
```

Вывод: как мы можем видеть из результатов, у всех комбинаций примерно одинаковая точность, но лучше всего с задачей справилась комбинация классификаторов - “Naive Bayes” + “K neighbours”.