МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ

(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)

Институт №8 «Информационные технологии и прикладная математика»

Кафедра 810 «Информационные технологии в моделировании и управлении»

**Лабораторная работа №3**

**по курсу «Интеллектуальный анализ данных»**

**Использование bootstrapping для оценки смещения и разброса алгоритмов машинного обучения. Бэггинг.**

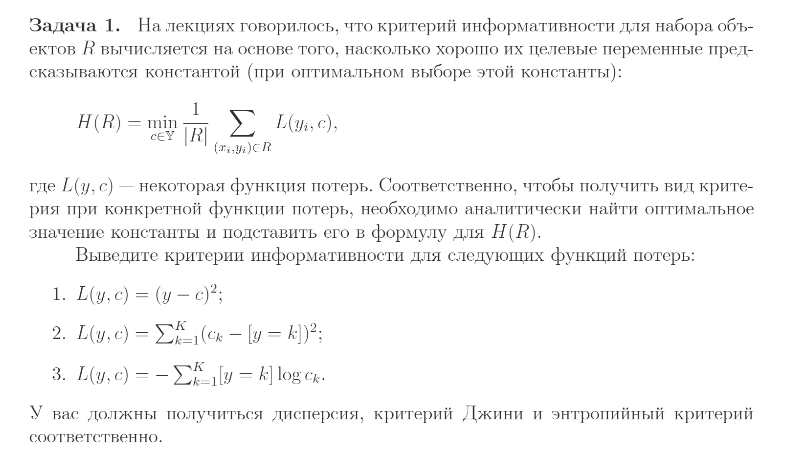
Выполнил: А.С.Бобряков

Группа: М8O-103М-19

Преподаватель: К.К.Абгарян

Москва, 2020

**Задание:**



**Решение:**

1)

Минимум в этом выражении будет достигаться на среднем значении целевой переменной. Значит, критерий можно переписать в следующем виде:

Получили, что информативность вершины измеряется ее дисперсией – чем ниже разброс целевой переменной, тем лучше вершина.

2)

Оптимальный вектор вероятностей состоит из долей классов pk, подставим его в H(R):

, ,

Получили критерий Джини.

3)

Для учета ограничения, что необходимо искать минимум лагранжиана:

Дифференцируем:

Выражаем и суммируем равенства по k:

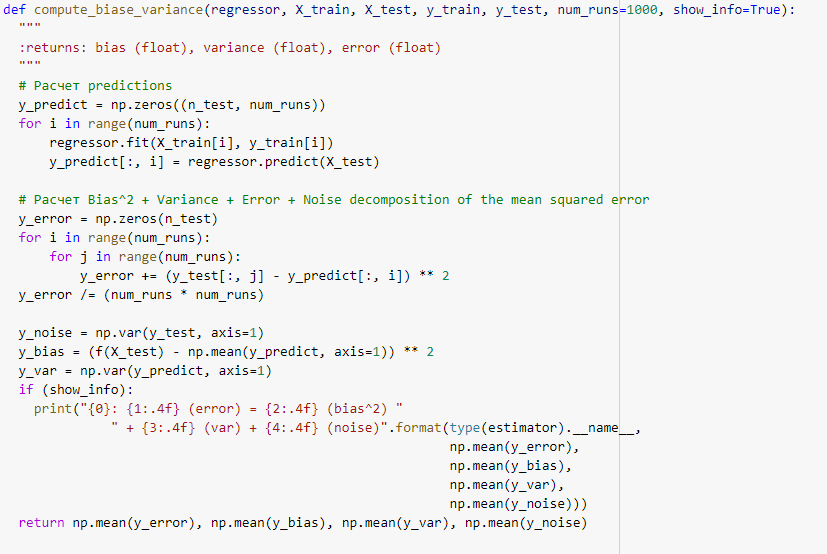
То есть то есть минимум достигается при , тогда критерий:

**Задание лабораторной работы:**

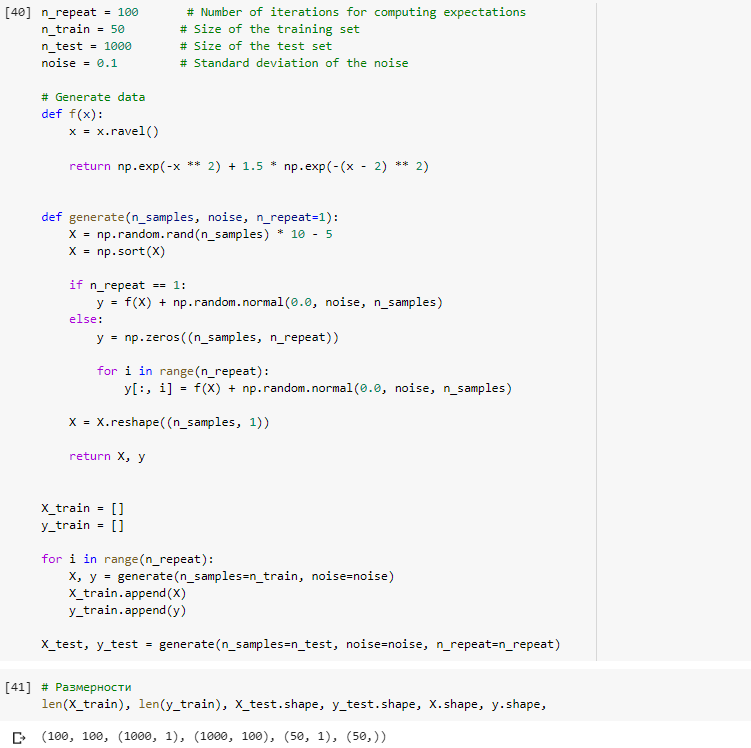
Реализовать алгоритм оценки смещения и разброса метода машинного обучения. Выполнить визуализацию предсказаний базовых алгоритмов бэггинга. Исследовать зависимость смещения и разбрася от гиперпаратметров алгоритма машинного обучения.

Ход выполнения лабораторной работы

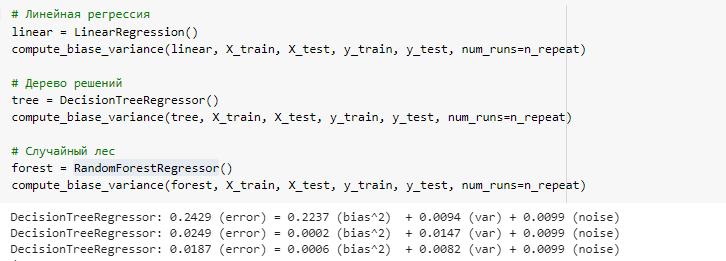
Функция расчета error, bias, var, noise:



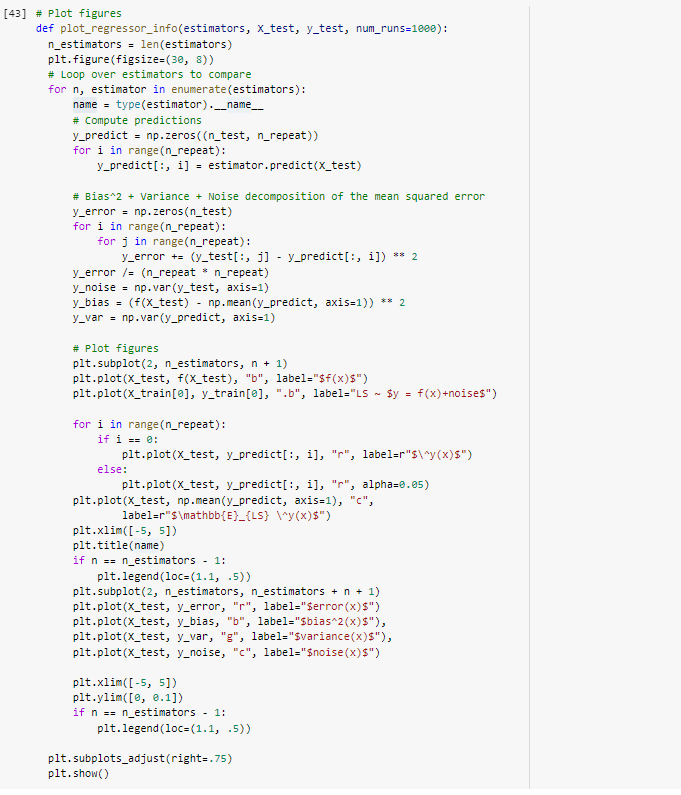
Генерируем данные bootstrap:



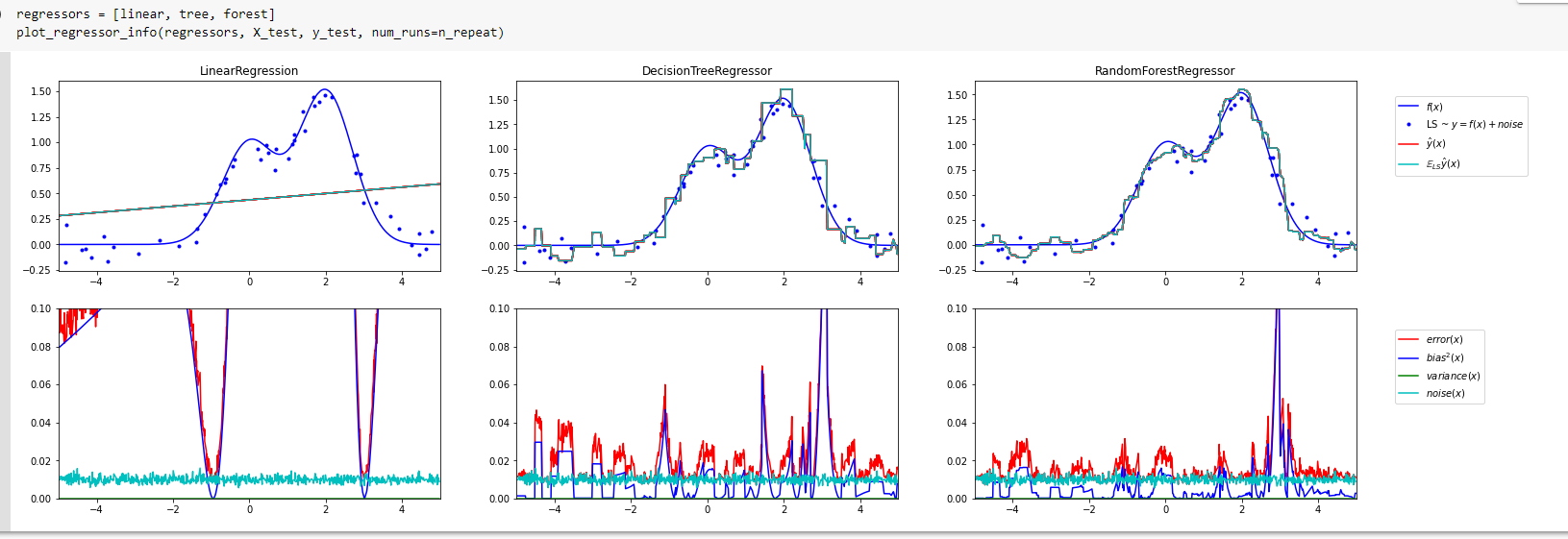
Строим модели:



Функция построения графиков для error, bias, var:



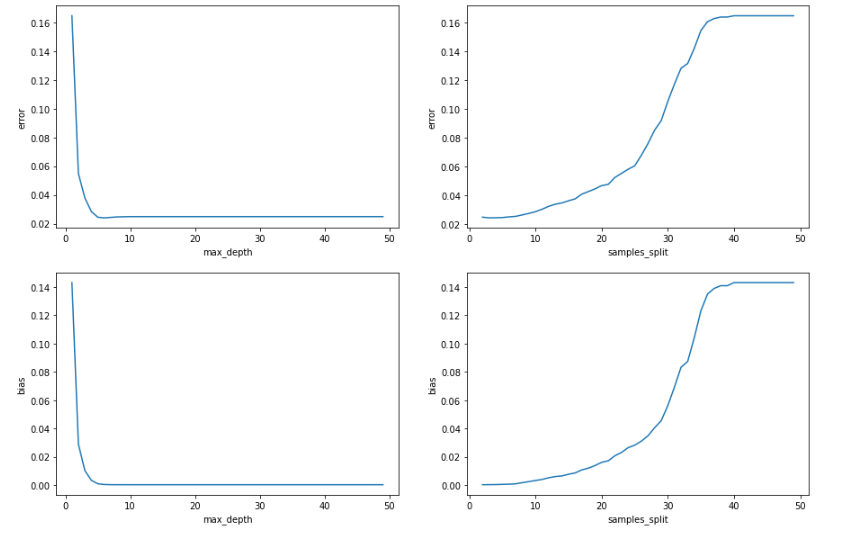
Строим графики для всех моделей:



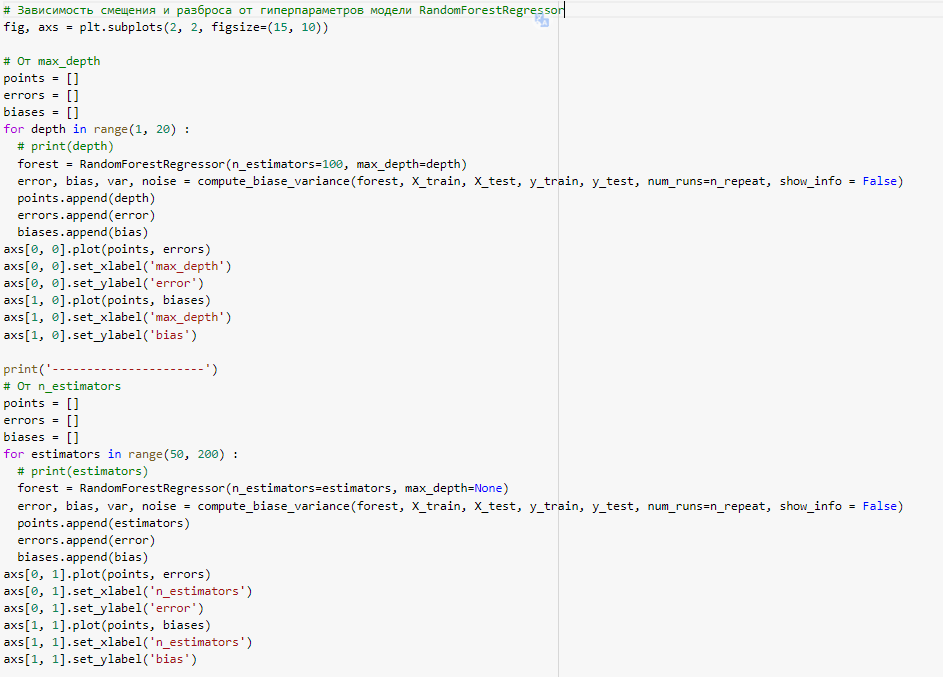
Графики зависимости error, bias от гиперпараметров моделей:

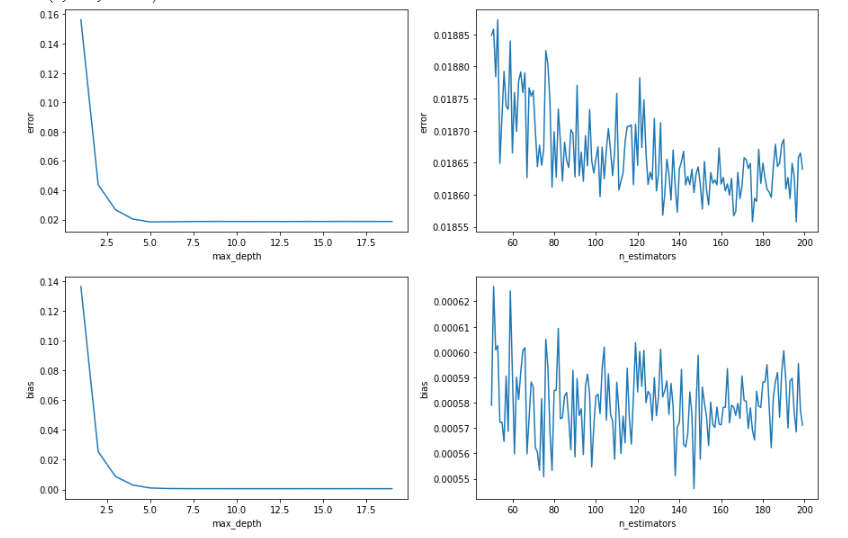
DecisionTreeRegressor





RandomForestRegressor





Выводы

В ходе лабораторной работы были изучен метод использования bootstrap, оценены смещения и разбросы моделей машинного обучения, проведена визуализация изученного материала, прослежена зависимость показателей качества моделей от гиперпараметров.