

МИНОБРНАУКИ РОССИИ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)
Кафедра МОЭВМ

ОТЧЕТ
по лабораторной работе №7
по дисциплине «Машинное обучение»
Тема: Классификация (Байесовские методы, деревья)

Студент гр. 8304

Холковский К.В

Преподаватель

Жангиров Т. Р.

Санкт-Петербург

2021

Цель работы

Ознакомиться с методами классификации модуля Sklearn.

Ход работы

Байесовские методы

1) Провели классификацию наблюдений наивным Байесовским методом.

Количество наблюдений, который были неправильно определены 5

Рис 1 – Количество наблюдений, который были неправильно определены

Таблица 1 – Атрибуты

Атрибуты	Смысл атрибута
class_count_	Количество обучающих выборок, наблюдаемых в каждом классе.
class_prior_	Вероятность каждого класса.
classes_	Метки классов, известные классификатору.
epsilon_	Фбсолютная аддитивная величина к отклонениям.
n_features_in_	Количество деталей, видимых во время посадки.
feature_names_in_	Названия особенностей, замеченных во время посадки. Определяется только тогда, когда X имеет имена функций, которые являются строками.
var_	Дисперсия каждого объекта по классу.
theta_	Среднее значение каждой функции по классу.

2) Вывели точность классификации

Точность классификации 0.96

Рис 2 - Точность классификации

3) Построили графики зависимости неправильно классифицированных наблюдений и точности классификации от размера выборки.

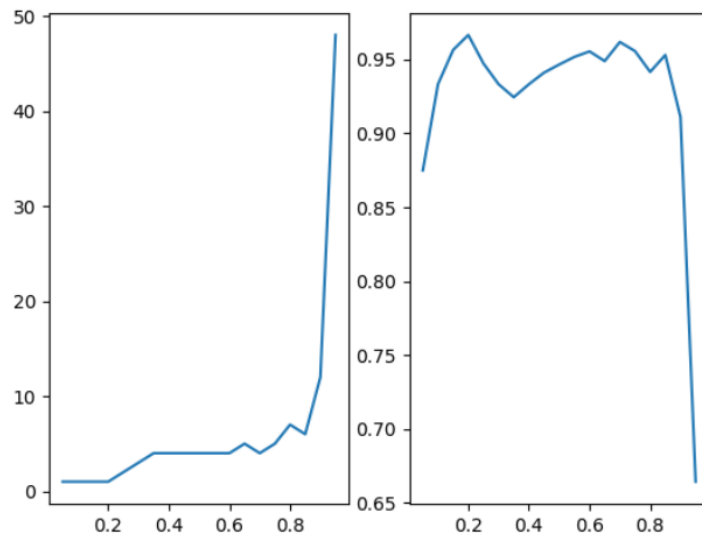


Рис 3— Графики зависимости неправильно классифицированных наблюдений и точности классификации от размера выборки GaussianNB

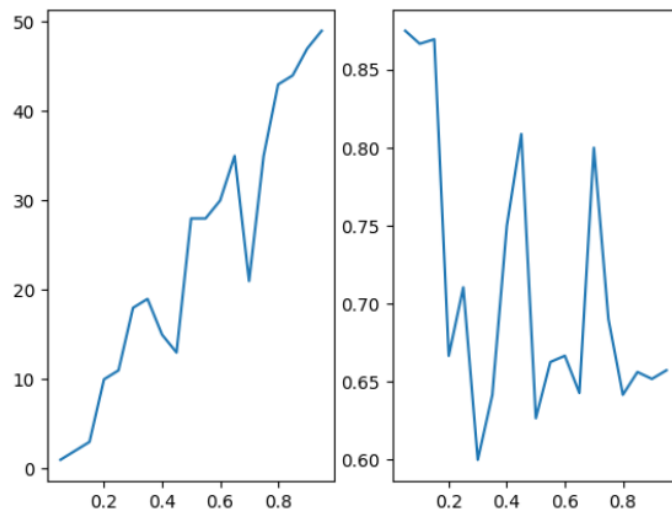


Рис 4— Графики зависимости неправильно классифицированных наблюдений и точности классификации от размера выборки MultinomialNB

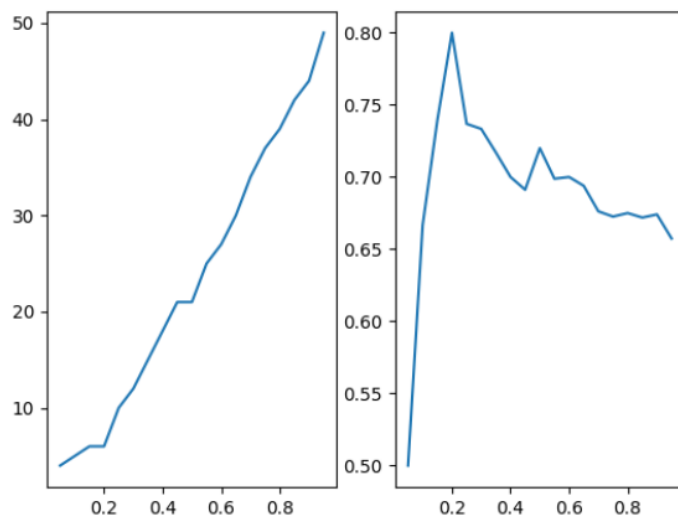


Рис 5— Графики зависимости неправильно классифицированных наблюдений и точности классификации от размера выборки ComplementNB

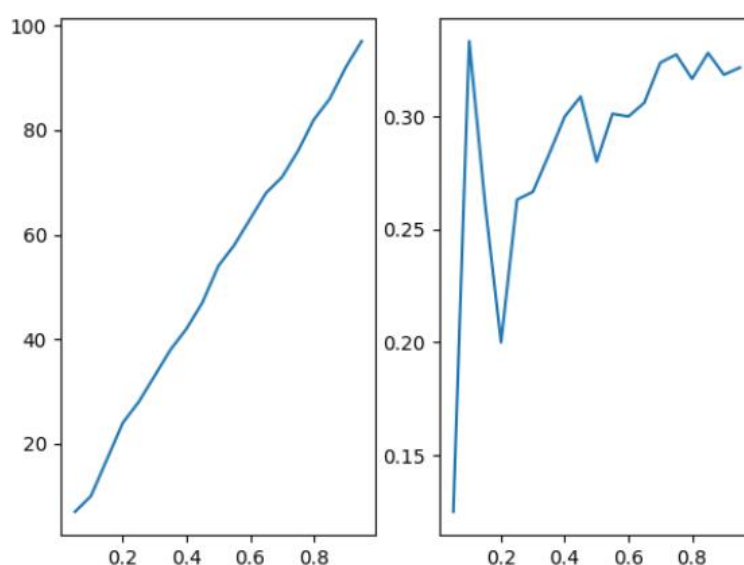


Рис 5— Графики зависимости неправильно классифицированных наблюдений и точности классификации от размера выборки BernoulliNB

Таблица 2 – Особенности методов

MultinomialNB	<p>Наивный байесовский классификатор для полиномиальных моделей.</p> <p>Полиномиальный наивный байесовский классификатор подходит для классификации с дискретными функциями (например, подсчетом слов для классификации текста). Полиномиальное распределение обычно требует целочисленного подсчета признаков. Однако на практике дробные подсчеты, такие как tf-idf, также могут работать.</p>
ComplementNB	<p>Наивный байесовский классификатор комплемента, описанный у Rennie et al. (2003).</p> <p>Дополнительный наивный байесовский классификатор был разработан для исправления «серьезных допущений», сделанных стандартным полиномиальным наивным байесовским классификатором. Он особенно подходит для несбалансированных наборов данных.</p>
BernoulliNB	<p>Наивный байесовский классификатор для многомерных моделей Бернулли.</p> <p>Как и MultinomialNB, этот классификатор подходит для дискретных данных. Разница в том, что в то время как MultinomialNB работает с подсчетом вхождений, BernoulliNB предназначен для двоичных / логических функций.</p>

Классифицирующие деревья

1) Провели классификацию на тех же данных

Количество наблюдений, который были неправильно определены 4
Точность классификации 1.0
Листьев: 5 ; Глубина: 4

Рис 6 – Результаты классификации

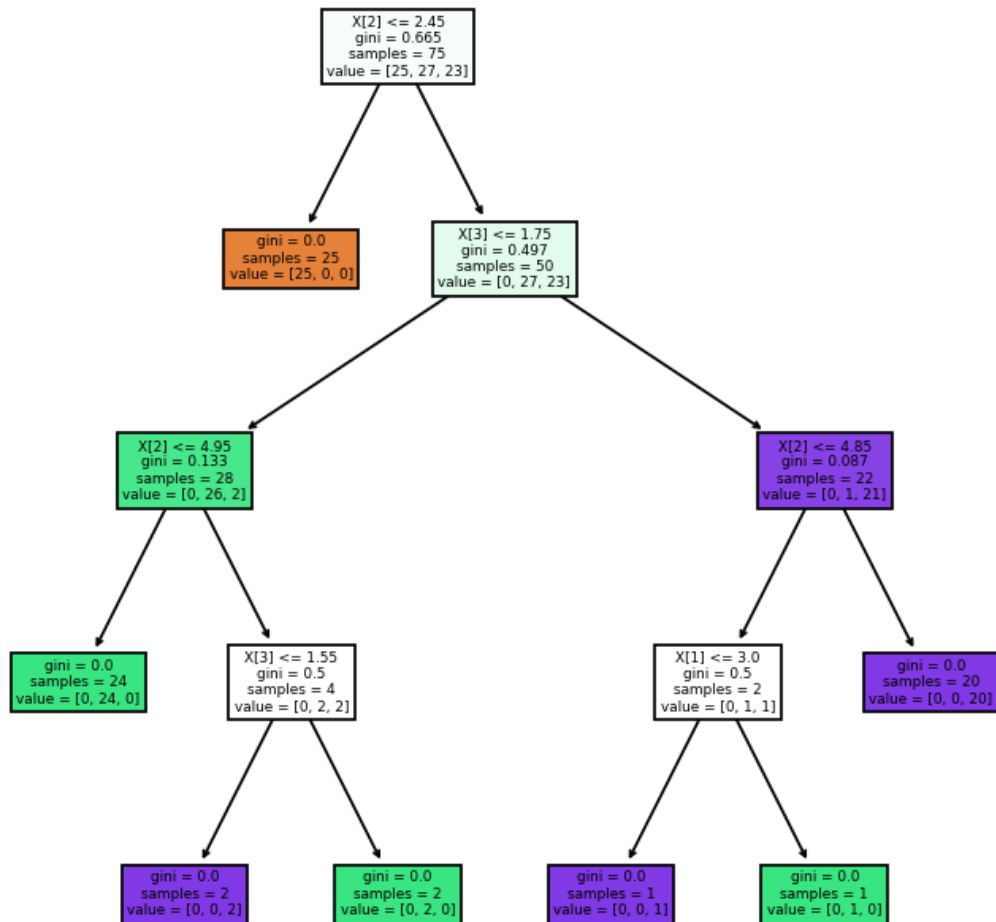


Рис 7 – Построенное дерево

2) Построили графики зависимости неправильно классифицированных наблюдений и точности классификации от размера выборки

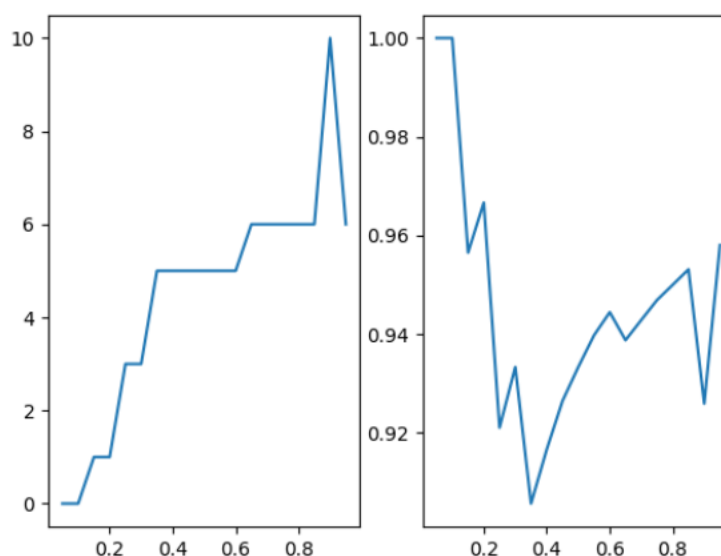


Рис 8 - Графики зависимости неправильно классифицированных наблюдений и точности классификации от размера выборки

3) Исследовали работу дерева при различных параметрах

Таблица 3 – Параметры дерева

Параметры	Смысл
criterion	Функция измерения качества раскола. Поддерживаемые критерии: «Джини» для примеси Джини и «энтропия» для получения информации.
splitter	Стратегия, используемая для выбора разделения на каждом узле. Поддерживаемые стратегии являются «лучшими» для выбора лучшего разбиения и «случайными» для выбора наилучшего случайного разбиения.
max_depth	Максимальная глубина дерева. Если None, то узлы расширяются до тех пор, пока все листья не станут чистыми или пока все листья не будут содержать менее <code>min_samples_split</code> выборок.
min_samples_split	Минимальное количество выборок, необходимое для разделения внутреннего узла.
min_samples_leaf	Минимальное количество выборок, которое требуется для конечного узла. Точка разделения на любой глубине будет учитываться только в том случае, если она оставляет по крайней мере обучающие образцы <code>min_samples_leaf</code> в каждой из левой и правой ветвей. Это может привести к сглаживанию модели, особенно при регрессии.

Вывод

Ознакомились с методами классификации модуля Sklearn.