**Практическая работа №6: «Наивный Байесовский классификатор»**

Оглавление

[Цель работы 1](#_Toc55298732)

[Задачи работы 1](#_Toc55298733)

[Перечень обеспечивающих средств 1](#_Toc55298734)

[Общие теоретические сведения 2](#_Toc55298735)

[**Логистическая регрессия** 2](#_Toc55298736)

[**Наивный байесовский классификатор** 2](#_Toc55298737)

[**Сравнение двух алгоритмов** 3](#_Toc55298738)

[Задание 4](#_Toc55298739)

[Требования к отчету 4](#_Toc55298740)

[Литература 4](#_Toc55298741)

## Цель работы

Получить практические навыки использования наивного Байесовского классификатора и логистической регрессии.

## Задачи работы

1. Научиться решать задачу классификации с помощью библиотеки sklearn, используя наивный Байесовский классификатор.
2. Научиться решать задачу классификации с помощью библиотеки sklearn, используя логистическую регрессию.
3. Научиться сравнивать результаты работы моделей классификации, используя F1-меру.

## Перечень обеспечивающих средств

1. ПК.
2. Учебно-методическая литература.
3. Задания для самостоятельного выполнения.

## Общие теоретические сведения

### **Логистическая регрессия**

Данные: , где ;

Задача: Найти такие значения , где *k* = 1, …, *K*, чтобы функция

,

где ,

аппроксимировала вероятность того, что .

Метрика производительности:

### **Наивный байесовский классификатор**

Данные: элементы ,

где

(размер набора данных), K – количество параметров, описывающих входные данные, M – количество классов.

Задача: Найти такую функцию , чтобы для всех *i*.

Вероятность того, что данный входной вектор *x* относится к данному классу

– априорные вероятности классов,

– функции правдоподобия.

«Наивность» байесовского классификатора:

Все параметры независимы, их порядок не имеет значения.

### **Сравнение двух алгоритмов**

Интересующие нас условные вероятности принадлежности к классу для имеющихся значений параметров:

* Наивный байесовский классификатор:

Вычисляются на основании вероятностей этих значений параметров (генеративный подход).

* Логистическая регрессия:

Вычисляются напрямую с помощью минимизации ошибки (дискриминативный подход).

## Задание

**Пояснение**

Для сохранения результатов данной работы вам понадобится два файла: doc/docx – для текста и ipynb – для кода. Назовите их одинаково: «*Фамилия* – задание 6».

**Часть 1**

* Обновите свой репозиторий, созданный в практической работе №1, из оригинального репозитория:

<https://github.com/mosalov/Notebook_For_AI_Main>.

**Часть 2**

* Откройте свой репозиторий в Binder (<https://mybinder.org/>).
* Откройте файл «task6.ipynb».
* Изучите, при необходимости – выполните повторно, приведённый в файле код.
* По аналогии с изученным выполните задание, приведённое в последней ячейке.
* Сохраните код в ipynb-файле. Необходимые пояснения опишите в своём docx/doc-файле.

## Требования к отчету

Оба файла (doc/docs и ipynb) загрузите в свой репозиторий, созданный в практическом задании №1 по пути: «Notebook\_For\_AI\_Main/2020 Осенний семестр/Практическое задание 6/» и сделайте пул-реквест.

## Литература

* <https://neurohive.io/ru/osnovy-data-science/kak-primenjat-teoremu-bajesa-dlja-reshenija-realnyh-zadach/>
* <https://habr.com/ru/post/170545/>
* <https://science.wikia.org/ru/wiki/Байесовская_вероятность>
* <https://dyakonov.org/2018/07/30/байесовский-подход/>
* <http://bazhenov.me/blog/2012/06/11/naive-bayes.html>
* <https://scikit-learn.org/stable/modules/naive_bayes.html>
* <https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html>
* https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1\_score.html