**Практическая работа № 3: «Машинное обучение. Способы машинного обучения. Типы решаемых задач. Метрики производительности»**

Оглавление

[Цель работы 1](#_Toc52359083)

[Задачи работы 1](#_Toc52359084)

[Перечень обеспечивающих средств 2](#_Toc52359085)

[Общие теоретические сведения 2](#_Toc52359086)

[**Регрессия. Метрики производительности.** 2](#_Toc52359087)

[**Классификация. Метрики производительности.** 2](#_Toc52359088)

[**Кластеризация. Метрики производительности.** 3](#_Toc52359089)

[Задание 4](#_Toc52359090)

[Требования к отчету 5](#_Toc52359091)

[Литература 5](#_Toc52359092)

## Цель работы

* Получить практические навыки расчёта метрик производительности для различных типов задач машинного обучения.
* На практическом примере разобрать различия между существующими метриками производительности для задачи классификации.

## Задачи работы

1. Используя результаты работы моделей машинного обучения, решающих задачи регрессии, классификации и кластеризации, научиться вычислять значения основных метрик производительности.
2. Реализовать программный код для вычисления компонентов матрицы путаницы.
3. Для задачи бинарной классицикации провести и проанализировать эксперименты по применимости различных метрик.

## Перечень обеспечивающих средств

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

## Общие теоретические сведения

### **Регрессия. Метрики производительности.**

*Среднеквадратичная ошибка:*

где - значение из данных, - результат работы модели.

*Средняя абсолютная ошибка:*

где - значение из данных, - результат работы модели.

### **Классификация. Метрики производительности.**

Матрица путаницы (ошибок):

|  |  |  |
| --- | --- | --- |
|  | y = 1 (выборка) | y = 2 (выборка) |
| y = 1 (модель) | TP | FP |
| y = 2 (модель) | FN | TN |

*accuracy* =

*precision* =

*recall* =

*F*1 =

### **Кластеризация. Метрики производительности.**

*Коэффициент силуэта*

Пусть для среднее расстояние между элементами кластера равно , а среднее расстояние до ближайшего кластера равно .

Тогда коэффициент силуэта для равен:

принимает значения из отрезка [-1, 1].

В качестве метрики используем среднее значение по всем .

## Задание

**Пояснение**

Для сохранения результатов данной работы вам понадобится два файла: doc/docx – для текста и ipynb – для кода. Назовите их одинаково: «*Фамилия* – задание 3».

**Часть 1**

* Некая модель, решающая задачу регрессии с помощью обучения с учителем, вернула следующие значения:

|  |  |  |
| --- | --- | --- |
| Описание объекта | Ожидаемый результат | Результат модели |
| 1, 2, 3 | 0 | -1 |
| 3, 5, 7 | 1 | 0 |
| 0, 0, 0 | 5 | 1 |
| 2, 8, 1 | 100 | 50 |

* Вычислите значения двух метрик регрессии для этой модели: среднеквадратичную ошибку и среднюю абсолютную ошибку. Сохраните результат в своём docx/doc-файле.

**Часть 2**

* Некая модель, решающая задачу бинарной классификации с помощью обучения с учителем, вернула следующие значения:

|  |  |  |
| --- | --- | --- |
| Описание объекта | Ожидаемый результат | Результат модели |
| 1, 2, 3 | 0 | 0 |
| 3, 5, 7 | 0 | 1 |
| 0, 0, 0 | 1 | 0 |
| 2, 8, 1 | 1 | 1 |
| 4, 4, 4 | 1 | 0 |
| 3, 4, 6 | 1 | 1 |
| 7, 5, 2 | 1 | 0 |
| 8, 8, 6 | 1 | 1 |

* Вычислите значение следующих метрик классификации для обоих классов (0 и 1) этой модели: accuracy, precision, recall и F1. Сохраните результат в своём docx/doc-файле.

**Часть 3**

* Некая модель, решающая задачу кластеризации с помощью обучения без учителя, вернула следующие значения (для двух классов):

|  |  |
| --- | --- |
| Описание объекта | Результат модели |
| 1, 2, 3 | 1 |
| 3, 5, 7 | 0 |
| 0, 0, 0 | 0 |
| 2, 8, 1 | 1 |

* Вычислите значение метрики кластеризации для этой модели – коэффициент силуэта – для каждой из записей и их среднее значение. При расчете используйте евклидово расстояние между объектами. Сохраните результат в своём docx/doc-файле.

**Часть 4**

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

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

**Часть 5**

* Откройте свой репозиторий в Binder (<https://mybinder.org/>).
* Откройте файл «task3.ipynb».
* Используйте свою фамилию для инициализации генератора случайных чисел, используя код в файле в качестве примера.
* Напишите свой код в соответствиями с инструкциями, 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 3/» и сделайте пул-реквест.

## Литература

* <https://ru.wikipedia.org/wiki/Машинное_обучение>
* <https://habr.com/ru/company/ods/blog/328372/>
* <https://ru.qwe.wiki/wiki/Silhouette_(clustering)>