

## Практическая работа №3

### Машинное обучение. Способы машинного обучения. Типы решаемых задач. Метрики производительности.

#### Оглавление

Цель работы .....	1
Задачи работы .....	1
Перечень обеспечивающих средств.....	2
Общие теоретические сведения .....	2
Регрессия. Метрики производительности.....	2
Классификация. Метрики производительности.....	2
Кластеризация. Метрики производительности.....	3
Задание .....	4
Требования к отчету .....	5
Литература .....	5

#### **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 классификации 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:*

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

где  $y_i$  - значение из данных,  $\hat{y}_i$  - результат работы модели.

*Средняя абсолютная ошибка:*

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

где  $y_i$  - значение из данных,  $\hat{y}_i$  - 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

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN}$$

$$precision = \frac{TP}{TP+FP}$$

$$recall = \frac{TP}{TP+FN}$$

$$F1 = 2 \frac{precision * recall}{precision + recall}$$

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

#### *Коэффициент силуэта*

Пусть для  $x_i$  среднее расстояние между элементами кластера равно  $a_i$ , а среднее расстояние до ближайшего кластера равно  $b_i$ .

Тогда коэффициент силуэта для  $x_i$  равен:

$$S(x_i) = \frac{b_i - a_i}{\max(a_i, b_i)}$$

$S(x_i)$  принимает значения из отрезка  $[-1, 1]$ .

В качестве метрики используем среднее значение  $S(x_i)$  по всем  $x_i$ .

## 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](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 соответствии 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/2021 Осенний семестр/Практическое задание 3/» и сделайте пул-реквест.

#### **Литература**

- [https://ru.wikipedia.org/wiki/Машинное\\_обучение](https://ru.wikipedia.org/wiki/Машинное_обучение)
- <https://habr.com/ru/company/ods/blog/328372/>
- [https://ru.qwe.wiki/wiki/Silhouette\\_\(clustering\)](https://ru.qwe.wiki/wiki/Silhouette_(clustering))