



Министерство науки и высшего образования Российской Федерации  
федеральное государственное бюджетное образовательное учреждение  
высшего профессионального образования  
«Московский государственный технический университет имени  
Н.Э. Баумана (национальный исследовательский университет)»  
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ «Информатика и системы управления»  
КАФЕДРА «Компьютерные системы и сети (ИУ-6)»

## ОТЧЕТ О ВЫПОЛНЕНИИ ЛАБОРАТОРНОЙ РАБОТЫ по дисциплине «Поддержка принятия решений в системах мониторинга»

Студент:	Козлов Владимир Михайлович
Группа:	ИУ6-13М
Тип задания:	лабораторная работа
Тема:	Выявление логических закономерностей по данным мониторинга

Студент

\_\_\_\_\_  
подпись, дата

Козлов В.М.

\_\_\_\_\_  
Фамилия, И.О.

Преподаватель

\_\_\_\_\_  
подпись, дата

\_\_\_\_\_  
Фамилия, И.О.

Москва, 2024

## Содержание

Цель .....	3
Задание .....	3
1   Выполнение работы .....	4
1.1   Признаки и классы. ....	4
2   Ответы на вопросы .....	7
3   Вывод .....	9

## Цель

Изучение способов выявления закономерностей в разнородных данных.

## Задание

Имеются два класса изображений лиц людей (смотри рисунок 1). Найти закономерности группирования этих изображений и определить, чем лица разных классов отличаются друг от друга и что объединяет лица одного класса.

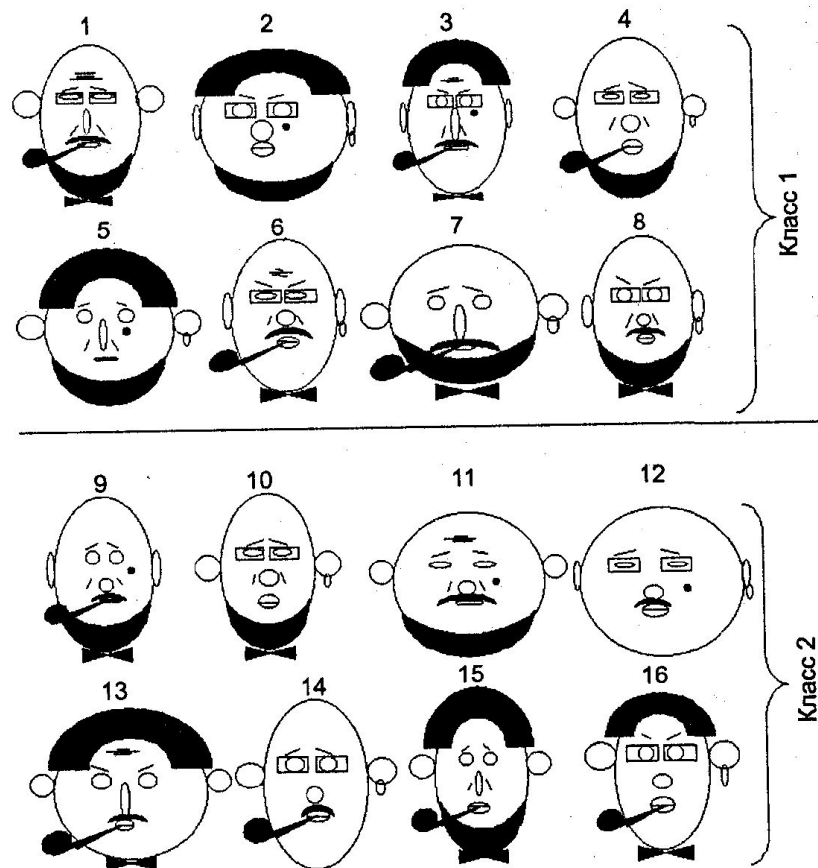


Рис. 1: 2 класса лиц

# 1 Выполнение работы

## 1.1 Признаки и классы

Были выделены следующие признаки:

1. (голова): круглая - 1, овальная - 0;
2. (уши): оттопыренные - 1, прижатые - 0;
3. (нос): круглый - 1, длинный - 0;
4. (глаза): круглые - 1, узкие - 0;
5. (лоб): с морщинами - 1, без морщин - 0;
6. (складка): носогубная складка есть - 1, носогубной складки нет - 0;
7. (губы): толстые - 1, тонкие - 0;
8. (волосы): есть - 1, нет - 0;
9. (усы): есть - 1, нет - 0;
10. (борода): есть - 1, нет - 0;
11. (очки): есть - 1, нет - 0;
12. (родинка): родинка на щеке есть - 1, родинки на щеке нет - 0;
13. (бабочка): есть - 1, нет - 0;
14. (брови): подняты кверху - 1, опущены книзу - 0;
15. (серьга): есть - 1, нет - 0;
16. (трубка): курительная трубка есть - 1, нет -
- 17.

Так как выделение закономерностей выполнено на языке Python, то и данные заносятся на нём. На следующем листинге можно увидеть массив элементов с кортежем из массива признаков и класса, к которому принадлежит элемент.

Листинг 1: Признаки и классы

---

```
1 Data = [  
2     ([0,1,0,0,1,1,0,0,1,1,1,0,1,1,0,1], 1),  
3     ([1,0,1,1,0,0,1,1,0,1,1,1,0,0,1,0], 1),  
4     ([0,0,0,1,1,1,0,1,1,0,1,1,1,0,0,1], 1),  
5     ([0,1,1,0,0,1,1,0,0,1,1,0,0,1,1,1], 1),
```

```

6 ([1,1,0,1,0,1,0,1,0,1,0,1,1,0], 1),
7 ([0,0,1,0,1,1,1,0,1,0,1,0,1,1], 1),
8 ([1,1,0,1,0,0,0,0,1,1,0,0,1,1,1], 1),
9 ([0,0,1,1,0,1,1,0,1,1,1,0,1,0,1,0], 1),
10 ([0,0,1,1,0,1,0,0,1,1,0,1,1,1,0,1], 2),
11 ([0,1,1,0,0,1,1,0,0,1,1,0,1,1,1,0], 2),
12 ([1,1,1,0,1,1,0,0,1,1,0,1,0,1,0,0], 2),
13 ([1,0,1,0,1,0,1,0,1,0,1,1,0,1,1,0], 2),
14 ([1,1,0,1,1,0,1,1,1,0,0,0,1,0,0,1], 2),
15 ([0,1,1,1,0,0,1,0,1,0,1,0,0,1,1,1], 2),
16 ([0,1,0,1,0,1,1,1,0,1,0,0,1,1,0,1], 2),
17 ([0,1,1,1,0,0,1,1,0,0,1,0,1,0,1,1], 2)
18 ]

```

Для решения используется наивный байесовский классификатор. На следующем листинге представлен код классификатора.

#### Листинг 2: Код классификатора

```

1 import numpy as np
2
3 class NaiveBayesG:
4     def fit(self, X, y):
5         self.classes, classesVolumes = np.unique(y, return_counts=True)
6         self.priories = classesVolumes / len(y)
7
8         self.X_mus = np.array([np.mean(X[y == c], axis=0) for c in self.classes])
9         self.X_sigmas = np.array([np.std(X[y == c], axis=0) for c in self.classes])
10
11     def gaussBayesFormula(self, x, mu, sigma):
12         return (1 / (np.sqrt(2 * np.pi) * sigma)) * np.exp(-((x - mu) ** 2 / (2 * sigma **
13             2)))
14
15     def predict(self, X):
16         possibilities = np.array([self.gaussBayesFormula(x, self.X_mus, self.X_sigmas) for x
17             in X])
18         posteriors = self.priories * np.prod(possibilities, axis=2)
19         return np.array([self.classes[pred] for pred in np.argmax(posteriors, axis=1)])

```

Для проверки адекватности проводилась следующим кодом:

#### Листинг 3: Проверка адекватности

```

1 import numpy as np
2 import data_original as data
3 from bayes import NaiveBayesG
4
5 Data = np.array([np.array(d[0]) for d in data.Data])
6 Classes = np.array([d[1] for d in data.Data])

```

```

7
8 train = [0,1,2,4,8,9,10,12]
9 test = list(range(len(Data)))
10 for i in train:
11     test.remove(i)
12
13 X_train = np.array([np.array(Data[i]) for i in train])
14 y_train = np.array([Classes[i] for i in train])
15
16 X_test = np.array([np.array(Data[i]) for i in test])
17 y_test = np.array([Classes[i] for i in test])
18
19 bayes = NaiveBayesG()
20 bayes.fit(X_train, y_train)
21
22 man_y_pred = bayes.predict(X_test)
23 print("Ошибка в%d случаев из%d" % ((y_test != man_y_pred).sum(), X_test.shape[0]))

```

---

**Результат:** Ошибка в 5 случаев из 8

**Вывод:** В силу малого количества данных байесовский классификатор не даёт хороших результатов. Также формула не сработает корректно если в классе у всех элементов совпадает признак так как в таком случае он бессмысленен.

## 2 Ответы на вопросы

### Вопросы

1. Что понимается под закономерностями в данных? Приведите промеры типовых закономерностей;
2. Поясните основные подходы выявления и анализа закономерностей внутри данных;
3. Укажите и поясните основные статистические методы для выявления закономерностей в данных;
4. Укажите интеллектуальные методы, применяемые для анализа больших данных;
5. С какой целью проводится кодирование информационных признаков?
6. Как можно определить логические закономерности в данных?
7. С какой целью проводится предварительная обработка данных при мониторинге? Что она включает?

### Ответы

1. Закономерности в данных — это повторяющиеся шаблоны или тенденции, которые можно обнаружить в наборе данных. Они помогают понять, как различные переменные взаимодействуют друг с другом и могут быть использованы для прогнозирования будущих значений или поведения.  
Вот несколько примеров типовых закономерностей:
  - (a) Линейные
  - (b) Сезонные
2.
  - (a) Статистический анализ - определение взаимосвязей исходя из статистики через основные характеристики и коэффициенты корреляции.
  - (b) Визуализация данных - визуализация данных в виде различных графиков или тепловых карт.
  - (c) Машинное обучение - разделение данных исходя из признаков путём классификации, регрессии или кластеризации.
  - (d) Анализ временных рядов - Используется для выявления закономерностей в данных, собранных во времени.
3. Основные статистические методы:
  - (a) Описательная статистика.

- (b) Корреляционный анализ.
  - (c) Регрессионный анализ.
  - (d) Кластерный анализ.
  - (e) Факторный анализ.
4. (a) машинное обучение;
- (b) дата майнинг;
- (c) нейросети;
- (d) имитационные модели;
- (e) предикативный и статистический анализ.
5. Чтобы можно было представить информацию в формате матрицы 0 и 1, и использовать её для математической обработки с помощью машинного обучения по средством написания кода. Или для простого более наглядного, упорядоченного представления информации.
6. Полученную информацию можно представить графически, использовать графы или закодировать и отсортировать в таблице. После чего либо ручным анализом (например деревья), либо с помощью компьютерных программ, выявляющим закономерности в закодированной информации.
7. Предварительная обработка данных при мониторинге проводится с целью улучшения качества данных и повышения точности анализа. Она включает в себя несколько ключевых этапов:
- (a) Очистка данных.
  - (b) Нормализация.
  - (c) Трансформация.
  - (d) Отбор признаков.
  - (e) Заполнение пропусков.
  - (f) Агрегация.



### 3 Вывод

в результате данной лабораторной работы был проведен системный анализ данных, изучены различные способы выявления закономерностей в данных и реализован один из них.