

Лабораторная работа №9 по дисциплине "Искусственный интеллект и машинное обучение"

Выполнил: студент 2-го курса Звездин Алексей Сергеевич

Группа: ПИЖ-6-о-22-1


Руководитель практики: Березина Виктория Андреевна, ассистент кафедры информационных систем и технологий института цифрового развития



Тема работы: Кластеризация

Цель работы: Научится производить кластерный анализ данных на основе метода К-средних

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

dataset = pd.read_csv('https://raw.githubusercontent.com/InternetHacker1123/bd_ai/main/laba9/list.data')
dataset.head()
```



	CustomerID	Genre	Age	Annual Income (k\$)	Spending Score (1-100)	
0	1	Male	24	25	57	
1	2	Male	22	26	78	
2	3	Female	23	27	34	
3	4	Female	21	28	82	
4	5	Female	29	29	55	

Next steps:  [View recommended plots](#)

```
X = dataset.iloc[:, [3, 4]].values
X
```



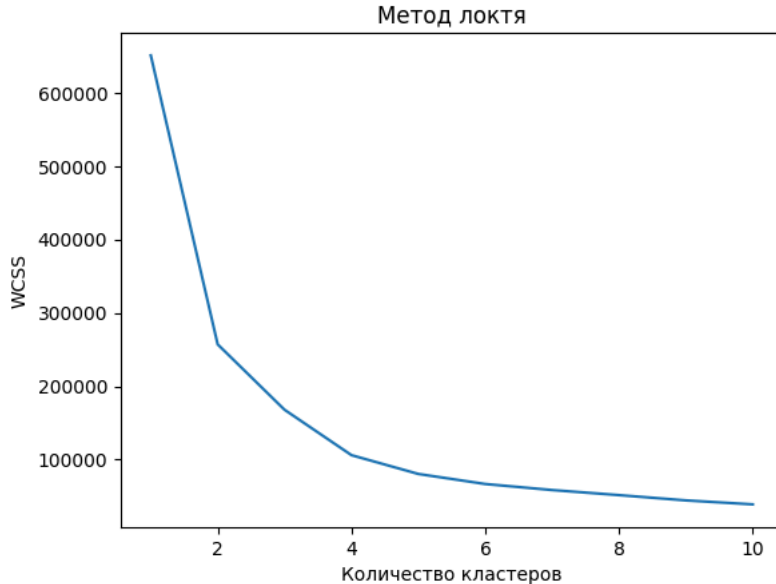
```
[179, 21],
[180, 94],
[181, 6],
[182, 83],
[183, 6],
[184, 78],
[185, 40],
[186, 88],
[187, 10],
[188, 98],
....]
```

```
from sklearn.cluster import KMeans
wcss = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters = i, init = 'k-means++', random_state = 42)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)
plt.plot(range(1, 11), wcss)
plt.title('Метод локтя')
plt.xlabel('Количество кластеров')
plt.ylabel('WCSS')
plt.show()
```

```

/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 1 in the future. To suppress this warning, please specify the value of `n_init` explicitly.
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 1 in the future. To suppress this warning, please specify the value of `n_init` explicitly.
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 1 in the future. To suppress this warning, please specify the value of `n_init` explicitly.
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 1 in the future. To suppress this warning, please specify the value of `n_init` explicitly.
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 1 in the future. To suppress this warning, please specify the value of `n_init` explicitly.
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 1 in the future. To suppress this warning, please specify the value of `n_init` explicitly.
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 1 in the future. To suppress this warning, please specify the value of `n_init` explicitly.
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 1 in the future. To suppress this warning, please specify the value of `n_init` explicitly.
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 1 in the future. To suppress this warning, please specify the value of `n_init` explicitly.
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 1 in the future. To suppress this warning, please specify the value of `n_init` explicitly.
warnings.warn(

```



```
kmeans = KMeans(n_clusters = 5, init = 'k-means++', random_state = 42)
y_kmeans = kmeans.fit_predict(X)
```

```

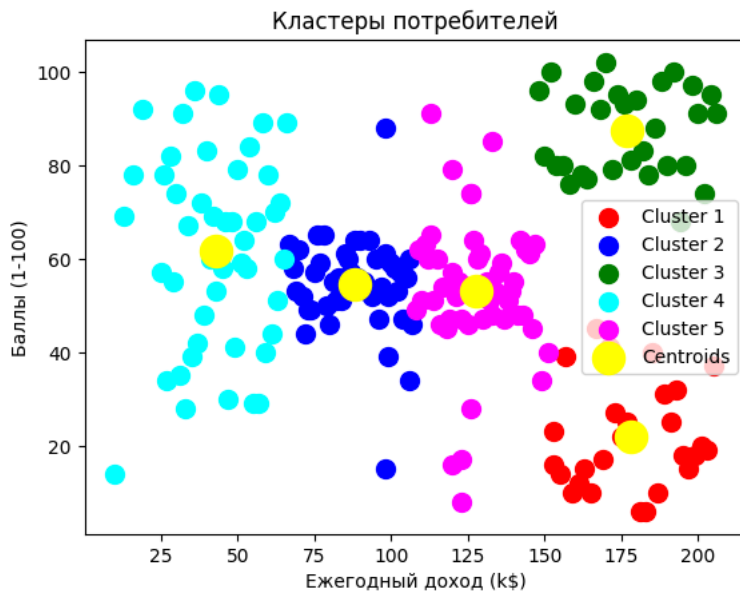
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 1 in the future. To suppress this warning, please specify the value of `n_init` explicitly.
warnings.warn(

```

```

plt.scatter(X[y_kmeans == 0, 0], X[y_kmeans == 0, 1], s = 100, c = 'red', label = 'Cluster 1')
plt.scatter(X[y_kmeans == 1, 0], X[y_kmeans == 1, 1], s = 100, c = 'blue', label = 'Cluster 2')
plt.scatter(X[y_kmeans == 2, 0], X[y_kmeans == 2, 1], s = 100, c = 'green', label = 'Cluster 3')
plt.scatter(X[y_kmeans == 3, 0], X[y_kmeans == 3, 1], s = 100, c = 'cyan', label = 'Cluster 4')
plt.scatter(X[y_kmeans == 4, 0], X[y_kmeans == 4, 1], s = 100, c = 'magenta', label = 'Cluster 5')
plt.scatter(kmeans.cluster_centers_[0, 0], kmeans.cluster_centers_[0, 1], s = 300, c = 'yellow', label = 'Centroids')
plt.title('Кластеры потребителей')
plt.xlabel('Ежегодный доход (k$)')
plt.ylabel('Баллы (1-100)')
```

```
plt.legend()
plt.show()
```



✓ Контрольные вопросы:

1. Что такое кластерный анализ?

Кластерный анализ - это метод машинного обучения, который используется для группировки схожих объектов внутри набора данных. Цель кластерного анализа состоит в том, чтобы разделить данные на группы (кластеры) таким образом, чтобы объекты внутри одного кластера были более похожи друг на друга, чем на объекты из других кластеров.

2. Перечислите известные методы кластерного анализа.

-Иерархическая кластеризация (Hierarchical Clustering) -Метод k-средних (k-means clustering) -DBSCAN (Density-Based Spatial Clustering of Applications with Noise) -Агломеративная кластеризация (Agglomerative Clustering) -Спектральная кластеризация (Spectral Clustering) -OPTICS (Ordering Points To Identify the Clustering Structure) -Mean Shift Clustering -Affinity Propagation -BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) -CLARA (Clustering Large Applications)

3. Перечислите классы и функции Python, которые задействованы при реализации кластерного анализа.

При реализации кластерного анализа в Python часто используются следующие классы и функции:

-sklearn.cluster

-scipy.cluster.hierarchy

-sklearn.metrics

-matplotlib.pyplot

-pandas и numpy

4. Опишите принцип определения оптимального количества кластеров.

Определение оптимального количества кластеров является важным этапом при проведении кластерного анализа, так как от правильного выбора числа кластеров зависит качество и интерпретируемость результатов.

5. Опишите принципиальные отличия методов регрессии, кластеризации и классификации.

Методы регрессии, кластеризации и классификации являются основными методами машинного обучения, но они имеют разные цели и подходы к обработке данных.

Метод регрессии используется для прогнозирования непрерывных значений. Основная задача регрессии - построить модель, которая предсказывает зависимую переменную на основе одной или нескольких независимых переменных. Регрессия использует статистические методы для анализа отношений между переменными и построения математической модели, которая может предсказывать значения зависимой переменной. Метод кластеризации используется для группировки объектов на основе их сходства. Основная задача кластеризации - найти внутри данных скрытые группы (кластеры), которые имеют схожие характеристики. Кластеризация не требует заранее определенных классов или меток, а основывается на сходстве объектов в пространстве признаков. Алгоритмы кластеризации стремятся минимизировать расстояние между объектами внутри кластера и максимизировать расстояние между кластерами. Метод классификации используется для прогнозирования принадлежности

объектов к определенным классам или категориям. Основная задача классификации - построить модель, которая способна отнести новый объект к одному из заранее определенных классов. Классификация использует обучение с учителем, где модель обучается на размеченных данных (с указанием класса каждого объекта). Модель классификации стремится построить границу принятия решений, которая разделяет объекты разных классов.