



Федеральное агентство по рыболовству
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Калининградский государственный технический университет»
(ФГБОУ ВО «КГТУ»)

УТВЕРЖДАЮ
Начальник УРОПС

Фонд оценочных средств
(приложение к рабочей программе дисциплины)
**«ПРОЕКТИРОВАНИЕ И РАЗРАБОТКА СИСТЕМ ИНТЕЛЛЕКТУАЛЬНОГО
АНАЛИЗА ДАННЫХ (ПРОЕКТНЫЙ ПРАКТИКУМ)»**

основной профессиональной образовательной программы магистратуры
по направлению подготовки

09.04.01 ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА

ИНСТИТУТ
РАЗРАБОТЧИК

цифровых технологий
кафедра прикладной математики и информационных технологий

1 РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Таблица 1 – Планируемые результаты обучения по дисциплине, соотнесенные с установленными индикаторами достижения компетенций

Код и наименование компетенции	Индикаторы достижения компетенции	Дисциплина	Результаты обучения (владения, умения и знания), соотнесенные с компетенциями/индикаторами достижения компетенции
ПК-3: Решает отдельные задачи разработки программного обеспечения в соответствии с техническим заданием	ПК-3.1: Решает отдельные задачи разработки программного обеспечения в соответствии с техническим заданием в области обработки больших данных	Проектирование и разработка систем интеллектуального анализа данных (проектный практикум)	<p><u>Знать:</u></p> <ul style="list-style-type: none"> - методы и инструментальные средства систем искусственного интеллекта, критерии их выбора и методы комплексирования в рамках создания интегрированных интеллектуальных систем анализа данных различного назначения; <p><u>Уметь:</u></p> <ul style="list-style-type: none"> - выбирать, применять и интегрировать методы и инструментальные средства систем интеллектуального анализа данных; - применять методы, языки и программные средства разработки программных компонентов систем интеллектуального анализа данных; <p><u>Владеть:</u></p> <ul style="list-style-type: none"> - навыками выбирать, адаптировать, разрабатывать и интегрировать программные компоненты систем ИАД с учетом установленных требований.

2 ПЕРЕЧЕНЬ ОЦЕНОЧНЫХ СРЕДСТВ И КРИТЕРИИ ОЦЕНИВАНИЯ

2.1 К оценочным средствам текущего контроля успеваемости относятся:

- тестовые задания открытого и закрытого типов.

2.2 К оценочным средствам для промежуточной аттестации относятся:

- курсовая работа по дисциплине;

- экзаменационные задания по дисциплине, представленные в виде тестовых заданий закрытого и открытого типов.

Промежуточная аттестация по дисциплине проводится в форме зачета, который выставляется по результатам прохождения всех видов текущего контроля успеваемости. При

необходимости тестовые задания закрытого и открытого типов могут быть использованы для проведения промежуточной аттестации.

2.3 Критерии оценки результатов освоения дисциплины

Универсальная система оценивания результатов обучения включает в себя системы оценок: 1) «отлично», «хорошо», «удовлетворительно», «неудовлетворительно»; 2) «зачтено», «не зачтено»; 3) 100 – балльную/процентную систему и правило перевода оценок в пятибалльную систему (табл. 2).

Таблица 2 – Система оценок и критерии выставления оценки

Система оценок Критерий	2	3	4	5
	0-40%	41-60%	61-80 %	81-100 %
	«неудовлетворительно»	«удовлетворительно»	«хорошо»	«отлично»
	«не зачтено»	«зачтено»		
1 Системность и полнота знаний в отношении изучаемых объектов	Обладает частичными и разрозненными знаниями, которые не может научно- корректно связывать между собой (только некоторые из которых может связывать между собой)	Обладает минимальным набором знаний, необходимым для системного взгляда на изучаемый объект	Обладает набором знаний, достаточным для системного взгляда на изучаемый объект	Обладает полной знаний и системным взглядом на изучаемый объект
2 Работа с информацией	Не в состоянии находить необходимую информацию, либо в состоянии находить отдельные фрагменты информации в рамках поставленной задачи	Может найти необходимую информацию в рамках поставленной задачи	Может найти, интерпретировать и систематизировать необходимую информацию в рамках поставленной задачи	Может найти, систематизировать необходимую информацию, а также выявить новые, дополнительные источники информации в рамках поставленной задачи
3.Научное осмысление изучаемого явления, процесса, объекта	Не может делать научно корректных выводов из имеющихся у него сведений, в состоянии проанализировать только некоторые из имеющихся у него сведений	В состоянии осуществлять научно корректный анализ предоставленной информации	В состоянии осуществлять систематический и научно корректный анализ предоставленной информации, вовлекает в исследование новые релевантные задачи данные	В состоянии осуществлять систематический и научно-корректный анализ предоставленной информации, вовлекает в исследование новые релевантные поставленной задаче

Система оценок Критерий	2	3	4	5
	0-40%	41-60%	61-80 %	81-100 %
	«неудовлетворительно»	«удовлетворительно»	«хорошо»	«отлично»
	«не зачтено»	«зачтено»		
				данные, предлагает новые курсы поставленной задачи
4. Освоение стандартных алгоритмов решения профессиональных задач	В состоянии решать только фрагменты поставленной задачи в соответствии с заданным алгоритмом, не освоил предложенный алгоритм, допускает ошибки	В состоянии решать поставленные задачи в соответствии с заданным алгоритмом	В состоянии решать поставленные задачи в соответствии с заданным алгоритмом, понимает основы предложенного алгоритма	Не только владеет алгоритмом и понимает его основы, но и предлагает новые решения в рамках поставленной задачи

2.4 Оценка проекта осуществляется поэтапно, этапы 3-4 предполагают открытую публичную защиту. При защите студент получает оценку «отлично», «хорошо», «удовлетворительно», «неудовлетворительно». Оценки выставляются согласно системе оценок и критериям их выставления, указанной в таблице 3.

Таблица 3 – Критерии оценивания

Показатель оценивания	Критерий оценивания	Оценка
Теоретическая и практическая ценность КП	Работа обладает новизной, имеет определенную теоретическую или практическую ценность	5
	Отдельные положения работы могут быть новыми и значимыми в теоретическом или практическом плане	4
	Работа представляет собой изложение известных фактов и не содержит рекомендации по их практическому использованию	3
	Полученные результаты или решение задачи не являются верными	2
Содержание проекта	Содержание полностью соответствует заявленной теме. Тема раскрыта полностью. Работа отличается логичностью. Выводы обоснованы	5
	Содержание работы соответствует заявленной теме, однако она раскрыта недостаточно обстоятельно. Работа выстроена достаточно логично	4
	Содержание работы не полностью соответствует заявленной теме, либо тема раскрыта недостаточно полно.	3
	Содержание работы не раскрывает заявленную тему.	2

Показатель оценивания	Критерий оценивания	Оценка
Использование источников	Общее количество используемых источников 10 и более, включая литературу на иностранных языках. Используется литература последних лет издания. Внутритекстовые ссылки и библиография оформлены в соответствии с ГОСТ	5
	Общее количество используемых источников не соответствует норме. Имеются погрешности в оформлении библиографического аппарата	4
	Количество используемых источников недостаточно или отсутствуют источники по теме работы. Используется литература давних лет издания. Имеются серьезные ошибки в библиографическом оформлении источников	3
	Изучено малое количество литературы. Нарушены правила внутритекстового цитирования, список литературы оформлен не в соответствии с действующим ГОСТ	2
Качество защиты	Студент демонстрирует хорошее знание вопроса, кратко и точно излагает свои мысли, умело ведет дискуссию.	5
	Студент владеет проблематикой и в целом правильно излагает свои мысли, однако ему не всегда удается аргументировать свою точку зрения при ответе на вопросы	4
	Студент затрудняется в кратком и четком изложении результатов своей работы.	3
	Студент плохо разбирается в теории вопроса. Не может изложить результаты своей работы.	2

3 ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

ПК-3: Решает отдельные задачи разработки программного обеспечения в соответствии с техническим заданием

Индикатор ПК-3.1: Решает отдельные задачи разработки программного обеспечения в соответствии с техническим заданием в области обработки больших данных.

Тестовые задания открытого типа:

1. В программном коде классификации рукописных символов (цифр)

```
n_samples = len(digits.images)
data = digits.images.reshape((n_samples, -1))

clf = svm.SVC(gamma=0.001)

X_train, X_test, y_train, y_test = train_test_split(
    data, digits.target, test_size=0.5, shuffle=False
)

clf.fit(X_train, y_train)

predicted = clf.predict(X_test)
```

используется классификатор: метод _____

Введите название метода (русск.)

Ответ: опорных векторов

2. В программном коде классификации рукописных символов (цифр)

```
n_samples = len(digits.images)
data = digits.images.reshape((n_samples, -1))

clf = svm.SVC(gamma=0.001)

X_train, X_test, y_train, y_test = train_test_split(
    data, digits.target, test_size=0.5, shuffle=False
)

clf.fit(X_train, y_train)

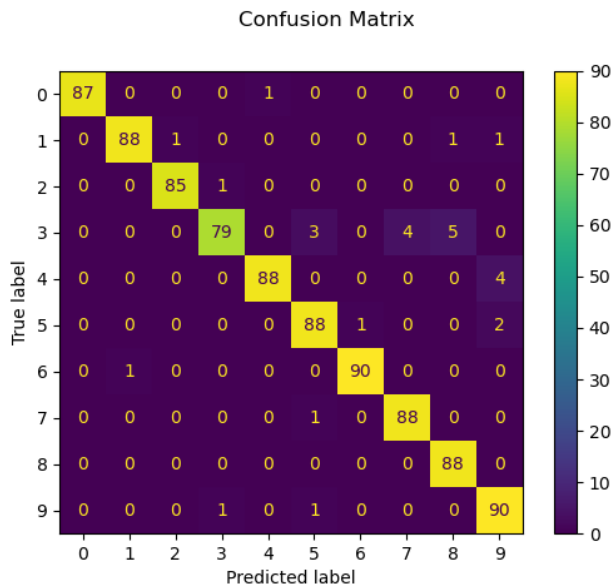
predicted = clf.predict(X_test)
```

тестовое и обучающее множества разделены в соотношении: _____

Введите соотношение в виде a:b

Ответ: 1:1 (или 50:50)

3. В результате обучения классификатора рукописных чисел методом SVM была получена матрица ошибок:



Количество неправильно классифицированных значений «9» равно: ____

Ответ: 2

4. В программном коде выполняется попиксельное квантование цветов изображения:

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.metrics import pairwise_distances_argmin
from sklearn.datasets import load_sample_image
from sklearn.utils import shuffle
from time import time

n_colors = 64
```

```
# Load the Summer Palace photo
china = load_sample_image("china.jpg")

# Convert to floats instead of the default 8 bits integer coding. Dividing by
# 255 is important so that plt.imshow behaves works well on float data (need to
# be in the range [0-1])
china = np.array(china, dtype=np.float64) / 255

# Load Image and transform to a 2D numpy array.
w, h, d = original_shape = tuple(china.shape)
assert d == 3
image_array = np.reshape(china, (w * h, d))

print("Fitting model on a small sub-sample of the data")
t0 = time()
image_array_sample = shuffle(image_array, random_state=0, n_samples=1_000)
kmeans = KMeans(n_clusters=n_colors, n_init="auto", random_state=0).fit(
    image_array_sample
)
print(f"done in {time() - t0:0.3f}s.")

# Get labels for all points
print("Predicting color indices on the full image (k-means)")
t0 = time()
labels = kmeans.predict(image_array)
print(f"done in {time() - t0:0.3f}s.")

codebook_random = shuffle(image_array, random_state=0, n_samples=n_colors)
print("Predicting color indices on the full image (random)")
t0 = time()
labels_random = pairwise_distances_argmin(codebook_random, image_array, axis=0)
print(f"done in {time() - t0:0.3f}s.")

def recreate_image(codebook, labels, w, h):
    """Recreate the (compressed) image from the code book & labels"""
    return codebook[labels].reshape(w, h, -1)
```

Число цветов конечного изображения, обработанного методом к-средних, равно: _____

Ответ: 64

5. В программном коде выполняется попиксельное квантование цветов изображения:

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.metrics import pairwise_distances_argmin
from sklearn.datasets import load_sample_image
from sklearn.utils import shuffle
from time import time

n_colors = 64

# Load the Summer Palace photo
china = load_sample_image("china.jpg")

# Convert to floats instead of the default 8 bits integer coding. Dividing by
```

```
# 255 is important so that plt.imshow behaves works well on float data (need to
# be in the range [0-1])
china = np.array(china, dtype=np.float64) / 255

# Load Image and transform to a 2D numpy array.
w, h, d = original_shape = tuple(china.shape)
assert d == 3
image_array = np.reshape(china, (w * h, d))

print("Fitting model on a small sub-sample of the data")
t0 = time()
image_array_sample = shuffle(image_array, random_state=0, n_samples=1_000)
kmeans = KMeans(n_clusters=n_colors, n_init="auto", random_state=0).fit(
    image_array_sample
)
print(f"done in {time() - t0:0.3f}s.")

# Get labels for all points
print("Predicting color indices on the full image (k-means)")
t0 = time()
labels = kmeans.predict(image_array)
print(f"done in {time() - t0:0.3f}s.")

codebook_random = shuffle(image_array, random_state=0, n_samples=n_colors)
print("Predicting color indices on the full image (random)")
t0 = time()
labels_random = pairwise_distances_argmin(codebook_random, image_array, axis=0)
print(f"done in {time() - t0:0.3f}s.")

def recreate_image(codebook, labels, w, h):
    """Recreate the (compressed) image from the code book & labels"""
    return codebook[labels].reshape(w, h, -1)
```

Функция, реализующая алгоритм к-средних, использует библиотеку Python: _____

Введите название библиотеки

Ответ: sklearn.cluster

6. Программный код реализует генерацию данных и кластеризацию методом DBSCAN:

```
import numpy as np
from sklearn.cluster import DBSCAN
from sklearn import metrics

from sklearn.datasets import make_blobs
from sklearn.preprocessing import StandardScaler

centers = [[1, 1], [-1, -1], [1, -1]]
X, labels_true = make_blobs(
    n_samples=750, centers=centers, cluster_std=0.4, random_state=0
)

X = StandardScaler().fit_transform(X)

db = DBSCAN(eps=0.3, min_samples=10).fit(X)
labels = db.labels_
```



```
# Number of clusters in labels, ignoring noise if present.
n_clusters_ = len(set(labels)) - (1 if -1 in labels else 0)
n_noise_ = list(labels).count(-1)
```

```
print("Estimated number of clusters: %d" % n_clusters_)
print("Estimated number of noise points: %d" % n_noise_)
```

В сгенерированном наборе данных число размеченных классов (кластеров) равно: _____

Ответ:3

7. Программный код реализует генерацию данных, кластеризацию методом DBSCAN и визуализацию результата:

```
import numpy as np
from sklearn.cluster import DBSCAN
from sklearn import metrics

from sklearn.datasets import make_blobs
from sklearn.preprocessing import StandardScaler

centers = [[1, 1], [-1, -1], [1, -1]]
X, labels_true = make_blobs(
    n_samples=750, centers=centers, cluster_std=0.4, random_state=0
)

X = StandardScaler().fit_transform(X)

db = DBSCAN(eps=0.3, min_samples=10).fit(X)
labels = db.labels_

# Number of clusters in labels, ignoring noise if present.
n_clusters_ = len(set(labels)) - (1 if -1 in labels else 0)
n_noise_ = list(labels).count(-1)

unique_labels = set(labels)
core_samples_mask = np.zeros_like(labels, dtype=bool)
core_samples_mask[db.core_sample_indices_] = True

colors = [plt.cm.Spectral(each) for each in np.linspace(0, 1, len(unique_labels))]
for k, col in zip(unique_labels, colors):
    if k == -1:
        # Black used for noise.
        col = [0, 0, 0, 1]

    class_member_mask = labels == k

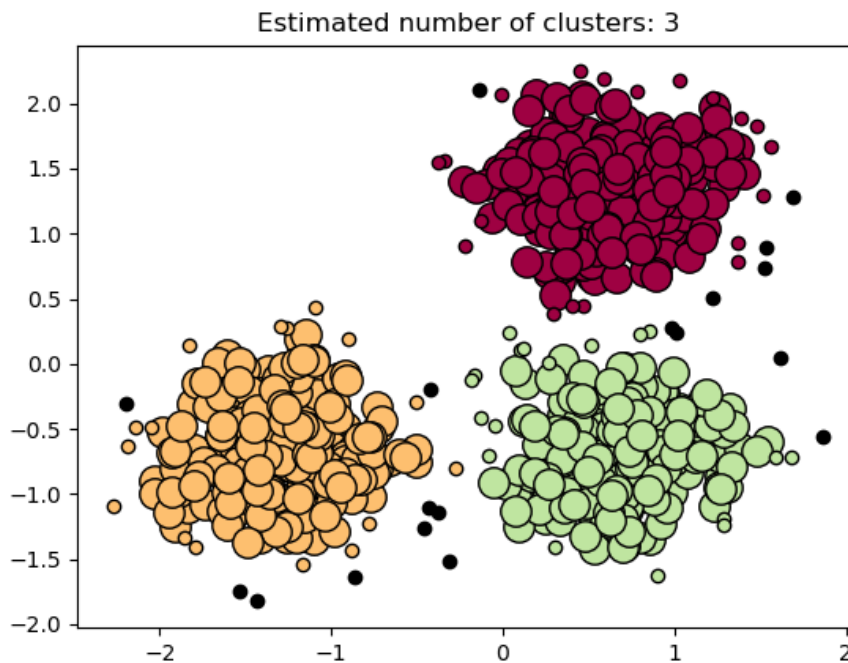
    xy = X[class_member_mask & core_samples_mask]
    plt.plot(
        xy[:, 0],
        xy[:, 1],
        "o",
        markerfacecolor=tuple(col),
        markeredgecolor="k",
        markersize=14,
    )

xy = X[class_member_mask & ~core_samples_mask]
```

```
plt.plot(
    xy[:, 0],
    xy[:, 1],
    "o",
    markerfacecolor=tuple(col),
    markeredgecolor="k",
    markersize=6,
)

plt.title(f"Estimated number of clusters: {n_clusters_}")
plt.show()
```

В результате визуализации:



классифицированные как шум точки имеют цвет: _____

Ответ: черный

8. Программный код реализует построение деревьев решений для набора данных «Ирисы Фишера»

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.datasets import load_iris

from sklearn.datasets import load_iris
from sklearn.tree import DecisionTreeClassifier
from sklearn.inspection import DecisionBoundaryDisplay

iris = load_iris()

# Parameters
n_classes = 3
plot_colors = "ryb"
plot_step = 0.02
```

```

for pairidx, pair in enumerate([[0, 1], [0, 2], [0, 3], [1, 2], [1, 3], [2, 3]]):
    # We only take the two corresponding features
    X = iris.data[:, pair]
    y = iris.target

    # Train
    clf = DecisionTreeClassifier().?????(X, y)

    # Plot the decision boundary
    ax = plt.subplot(2, 3, pairidx + 1)
    plt.tight_layout(h_pad=0.5, w_pad=0.5, pad=2.5)
    DecisionBoundaryDisplay.from_estimator(
        clf,
        X,
        cmap=plt.cm.RdYlBu,
        response_method="predict",
        ax=ax,
        xlabel=iris.feature_names[pair[0]],
        ylabel=iris.feature_names[pair[1]],
    )

    # Plot the training points
    for i, color in zip(range(n_classes), plot_colors):
        idx = np.where(y == i)
        plt.scatter(
            X[idx, 0],
            X[idx, 1],
            c=color,
            label=iris.target_names[i],
            cmap=plt.cm.RdYlBu,
            edgecolor="black",
            s=15,
        )

plt.suptitle("Decision surface of decision trees trained on pairs of features")
plt.legend(loc="lower right", borderpad=0, handletextpad=0)
_ = plt.axis("tight")

```

Вместо **????** требуется ввести метод: _____

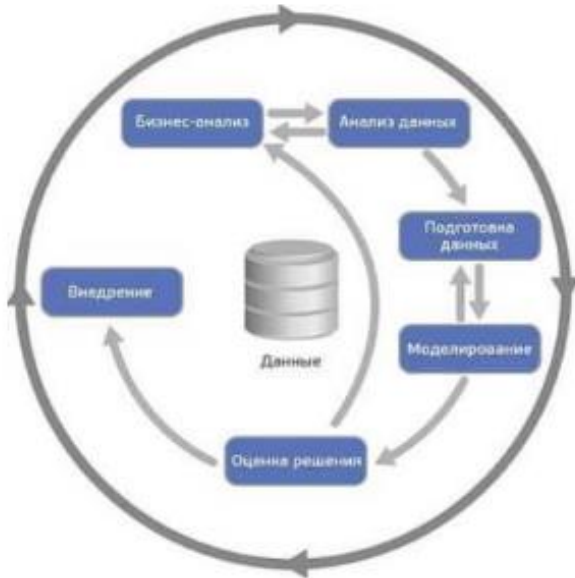
Регистр важен

Ответ: fit

9. _____ – это визуализатор, используемый для представления результатов иерархической кластеризации.

Ответ: дендрограмма

10. На рисунке



представлен межотраслевой стандартный процесс исследования данных: _____

Введите аббревиатуру (англ), регистр важен

Ответ: CRISP-DM

11. Класс алгоритмов, являющийся элегантной идеей по построению разделяющей поверхности, а также осуществляющий переход в новое пространство значительно дешевле, чем вычисление всех обучающие объектов в новом пространстве напрямую – это: _____

Введите аббревиатуру (англ.)

Ответ: SVM

12. Разность между желаемым и полученным на выходе сигналами – это _____ обучения нейронной сети

Ответ: ошибка

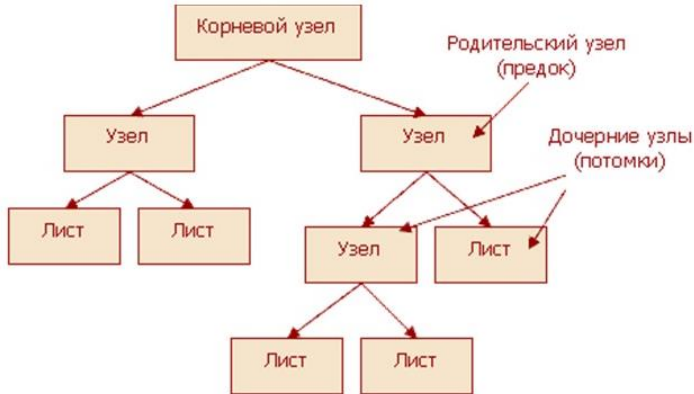
13. _____ нейронная сеть:

- каждый нейрон передает свой выходной сигнал остальным нейронам, в том числе и самому себе;
- все входные сигналы подаются всем нейронам;
- выходными сигналами сети могут быть все или некоторые выходные сигналы нейронов после нескольких тактов функционирования сети.

Введите вид архитектуры сети

Ответ: полносвязная

14. Рисунок



изображает: _____

Ответ: дерево решений

15. При обучении с учителем использование алгоритма обратного распространения ошибки может вызвать _____ сети, когда в процессе обучения весовые значения в результате коррекции становятся очень большими величинами и практически не меняются, а процесс обучения останавливается.

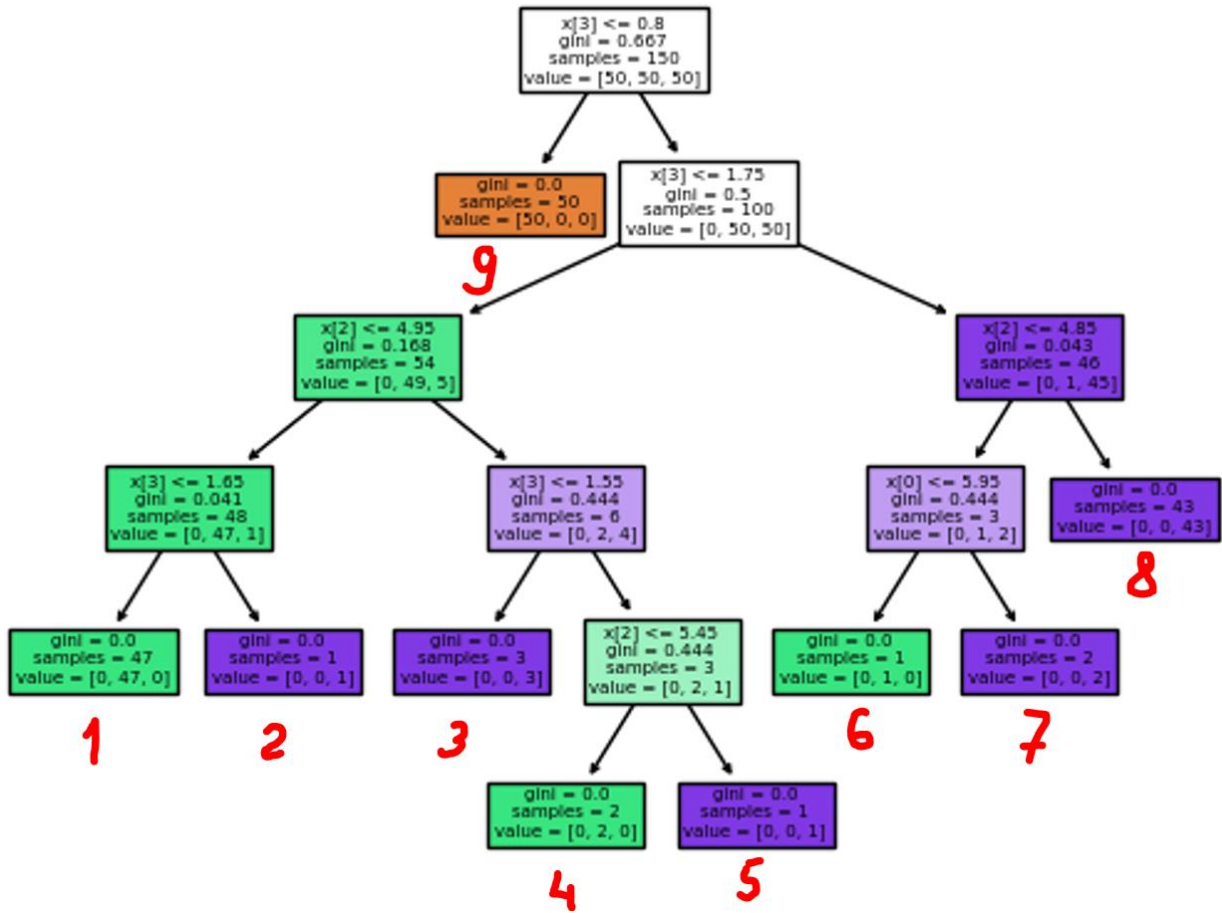
Введите название явления

Ответ: паралич

16. В результате выполнения кода

```
from sklearn.tree import plot_tree
plt.figure()
clf = DecisionTreeClassifier().fit(iris.data, iris.target)
plot_tree(clf, filled=True)
plt.title("Decision tree trained on all the iris features")
plt.show()
```

получено изображение дерева решений для классификатора набора данных «Ирисы Фишера», приведенное ниже



Классифицированный объект с параметрами $X=(0.1, 0.3, 0.8)$ соответствует листу с номером: _____

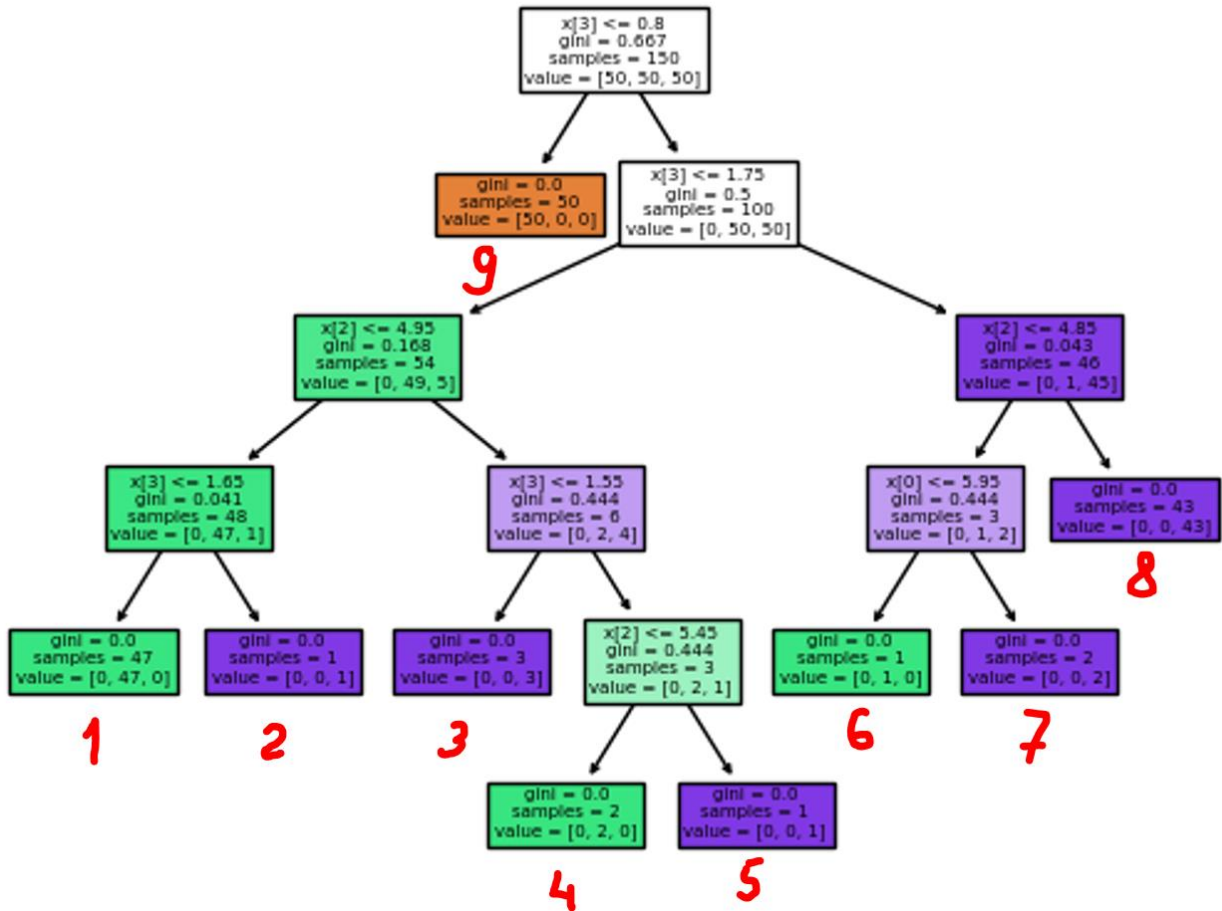
Ответ: 9

17. В результате выполнения кода

```
from sklearn.tree import plot_tree

plt.figure()
clf = DecisionTreeClassifier().fit(iris.data, iris.target)
plot_tree(clf, filled=True)
plt.title("Decision tree trained on all the iris features")
plt.show()
```

получено изображение дерева решений для классификатора набора данных «Ирисы Фишера», приведенное ниже:



К листу дерева с номером 8 приводит правило: _____

Введите последовательность условий через точку с запятой, без пробелов, разделитель рядов запятая, в виде:

x[m] знак_сравнения число

Допускается иная формулировка ответа, не искажающая его смысла

Ответ: x[3]>0,8;x[2]>4,85

18. В программном коде:

```

# Code source: Gael Varoquaux
# License: BSD 3 clause

import matplotlib.pyplot as plt
import numpy as np
from scipy.special import expit
from sklearn.linear_model import LinearRegression, LogisticRegression

# Generate a toy dataset, it's just a straight line with some Gaussian noise:
xmin, xmax = -5, 5
n_samples = 100
np.random.seed(0)
X = np.random.normal(size=n_samples)
y = (X > 0).astype(float)
X[X > 0] *= 4
X += 0.3 * np.random.normal(size=n_samples)
  
```

```

X = X[:, np.newaxis]

# Fit the classifier
clf = LogisticRegression(C=1e5)
clf.fit(X, y)

# and plot the result
plt.figure(1, figsize=(4, 3))
plt.clf()
plt.scatter(X.ravel(), y, label="example data", color="black", zorder=20)
X_test = np.linspace(-5, 10, 300)

loss = expit(X_test * clf.coef_ + clf.intercept_).ravel()
plt.plot(X_test, loss, label="Logistic Regression Model", color="red", linewidth=3)

ols = LinearRegression()
ols.fit(X, y)
plt.plot(
    X_test,
    ols.coef_ * X_test + ols.intercept_,
    label="Linear Regression Model",
    linewidth=1,
)
plt.axhline(0.5, color=".5")

plt.ylabel("y")
plt.xlabel("X")
plt.xticks(range(-5, 10))
plt.yticks([0, 0.5, 1])
plt.ylim(-0.25, 1.25)
plt.xlim(-4, 10)
plt.legend(
    loc="lower right",
    fontsize="small",
)
plt.tight_layout()
plt.show()

```

реализована классификация объектов с количеством классом: _____

Введите число

Ответ:2

19. В Logiном исполнении сценариев с формированием итоговых данных и передачей их в стороннюю систему без визуализации результатов на экране происходит в _____ режиме.

Ответ: пакетном

20. В Python в библиотеке sklearn многослойный перцептрон, обучаемый алгоритмом обратного распространения ошибки, реализует класс: _____

Регистр важен

Ответ: MLPClassifier

21. Статистический показатель, с помощью которого можно описывать характер изменения одной величины относительно изменения другой – это _____

Ответ: индекс Джини

22. В программном коде:

```
from sklearn.neural_network import MLPClassifier
X = [[0., 0.], [1., 1.]]
y = [0, 1]
clf = MLPClassifier(solver='lbfgs', alpha=1e-5,
                   hidden_layer_sizes=(5, 2), random_state=1)
clf.fit(X, y)
```

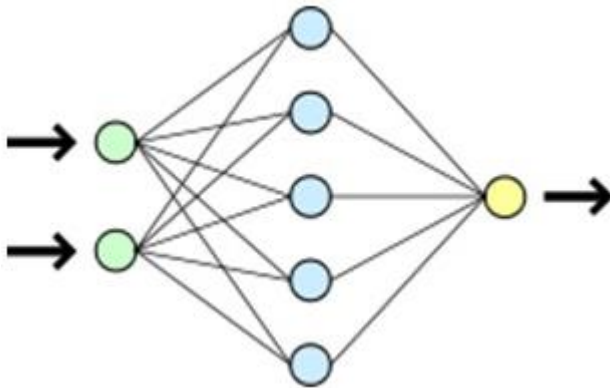
количество скрытых слоев нейронной сети равно: ____

Ответ:2

23. В Python после обучения нейронной сети MLPClassifier для разметки нового (тестового) набора данных используется метод:

Ответ: predict

32.



Количество скрытых слоев в нейронной сети на рисунке равно: ____

Ответ:1

24. В Logiном за публикацию собственных веб-сервисов на основе созданных пакетов отвечает компонент: _____

Введите название (англ.)

Ответ: Integrator

25. Архитектурный стиль взаимодействия компонентов распределённого приложения в сети, представляющий собой согласованный набор ограничений, учитываемых при проектировании распределённой гипермедиа-системы – это ____ - серви

Введите аббревиатуру (англ), регистр важен

Ответ: REST

Тестовые задания закрытого типа:

1. Суть алгоритма Expectation-Maximization:

1. выбор вектора параметров для максимизации функции правдоподобия, соответствующей семейству параметрических моделей.

2. максимизация математического ожидания многомерного нормального распределения с неизвестными параметрами.

3. присваивание ожидаемых значений скрытым переменным на основании текущих оценок параметров и переоценка этих параметров с учетом обновленных ожидаемых значений.

4. вычисление математического ожидания максимума функции правдоподобия, построенной на входной выборке.

2. Основная задача кластеризации:

1. поиск "скрытой структуры" данных

2. деление набор данных на predetermined классы

3. определение центра набора данных

4. упорядочивание объектов в статистически однородные группы

3. Для решения проблемы чувствительности функции расстояния к преобразованиям в данных необходимо преобразовать:

1. обучающую выборку так, чтобы оси совпадали с главными компонентами матрицы ковариации

2. тестовую выборку так, чтобы оси не совпадали с главными компонентами матрицы ковариации

3. обучающую выборку так, чтобы оси совпадали с второстепенными компонентами матрицы ковариации

4. тестовую выборку так, чтобы оси не совпадали с второстепенными компонентами матрицы ковариации

4. В методологии CRISP-DM **ЛИШНИМ** этапом будет:

1. понимание бизнес-целей

2. понимание данных

3. подготовка данных

4. обучение модели

5. моделирование

6. оценка

7. внедрение

5. Для преобразования многомерного пространства в пространство меньшей размерности и формирования малого количества признаков из большого количества признаков следует использовать алгоритм:

1. T-SNE

2. DBSCAN

3. PAM

4. CWM

6. В иерархической кластеризации существуют подходы:

Возможно несколько вариантов ответа

1. агломеративный

2. классификационный

3. дивизионный

4. разделительный

7. Кластеризация DBSCAN основана на:

1. отсутствии шумов в наборе данных

2. плотности объектов

3. выявлении граничных объектов

4. расширении кластеров

8. "Все переменные являются одинаково важными и статистически независимыми, т.е. значение одной переменной ничего не говорит о значении другой". Это свойства:

1. наивной байесовской классификации

2. атрибута расщепления

3. метода "к-ближайших соседей"

4. сети Кохонена

4 ТИПОВЫЕ ЗАДАНИЯ НА КОНТРОЛЬНУЮ РАБОТУ, КУРСОВУЮ РАБОТУ/КУРСОВОЙ ПРОЕКТ

4.1 Курсовой проект (КП) направлен на закрепление полученных теоретических знаний и приобретение умений и навыков в области проектирования систем интеллектуального анализа данных и машинного обучения.

КП может выполняться индивидуально, либо в составе проектной команды. В случае командной разработки студенты готовят общую пояснительную записку с указанием (распределением) ролей участников команды.

Примерное задание и темы КП ниже.

Задание: Разработка систем или элементов системы интеллектуального анализа данных, включая сбор, хранение и ETL-процессы.

План выполнения курсового проекта включает этапы, продолжающие содержание и документирование, приведенные в таблице.

№ этапа	Наименование работ по основным этапам НИР	Содержание этапа	Перечень документов
1	<i>Постановка задачи</i>	формулировка целей и задач проекта, анализ данных и источников, выбор методов и алгоритмов анализа	1. Техническое задание
2	<i>Проектирование и реализация - 1</i>	Итеративный процесс формулирования гипотез и идей для тестирования, проведение эксперимента, отработка методов и технологий, сбора данных.	1. Статья в сборник СНТК или аналогичной конференции 2. Презентация к СНТК
3	<i>Проектирование и реализация - 2</i>	Итеративный процесс создание минимально работоспособной версии продукта (MVP) готового для демонстрации и обсуждения	1. Пояснительная записка к проекту 2. Презентация к защите
4	<i>Оценка качества и документирование</i>		1. Презентация к защите для проектов, претендующих на оценку хорошо или отлично (на выбор) 2. Проект заявки на конкурс УМНИК 3. Заявление о государственной регистрации программы для ЭВМ или базы данных (с комплектом сопроводительных документов)

Типовые темы курсовых проектов:

Проектирование информационно-аналитической платформы анализа данных.

Проектирование программного комплекса анализа данных в рамках технологии DataMining.

Проектирование и программная реализация методов машинного обучения для решения бизнес-задачи.

К защите КП допускается обучающийся, выполнивший работу по утвержденной теме.

5 СВЕДЕНИЯ О ФОНДЕ ОЦЕНОЧНЫХ СРЕДСТВ И ЕГО СОГЛАСОВАНИИ

Фонд оценочных средств для аттестации по дисциплине «Проектирование и разработка систем интеллектуального анализа данных (проектный практикум)» представляет собой компонент основной профессиональной образовательной программы магистратуры по направлению подготовки 09.04.01 Информатика и вычислительная техника.

Фонд оценочных средств рассмотрен и одобрен на заседании методической комиссии института цифровых технологий (протокол № 2 от 26.04.2022 г.).

Фонд оценочных средств актуализирован. Изменения, дополнения рассмотрены и одобрены на заседании методической комиссии института цифровых технологий (протокол № 3 от 24.03.2023 г.).

Директор института



А.Б. Тристанов