**Методичка: ML0.1 intro**

1. Библиотека matplotlib используется в программировании и научных исследованиях для:

A) Визуализации данных +

B) Написания текстовых отчетов

C) Анализа синтаксической структуры текста

D) Работы с базами данных

E) Создания веб-приложений

2. В pandas возможен доступ к индексам следующими способами:

A) loc - используется для доступа по строковой метке +

B) iloc - используется для доступа по числовому значению (начиная от 0) +

C) loc - используется для доступа по числовому значению (начиная от 0)

D) iloc - используется для доступа по строковой метке

E) Обращение к названиям нужных столбцов (верно)

F) Методом index

**Методичка: ML0.2 simple analisys**

3. Библиотека pandas предоставляет следующие инструменты для работы с данными:

A) Методы для анализа и прогнозирования

B) Методы работы с базами данных SQL

C) Методы для чтения и записи данных из различных форматов файлов +

D) Функции для обработки и агрегации данных +

E) Инструменты для работы с пропущенными данными +

F) Работа с линейной алгеброй

4. Метод dropna в Pandas действует следующим образом:

A) Удаляет первую строку DataFrame.

B) Удаляет все строки с отсутствующими значениями из DataFrame. +

C) Удаляет последний столбец DataFrame.

D) Заполняет недостающие значения нулями в DataFrame. +

E) Переименовывает столбцы DataFrame.

F) Преобразует все значения в DataFrame в верхний регистр.

**Методичка: ML1.1 linear regression**

5 Коэффициент наклона (slope) в модели линейной регрессии можно интерпретировать как:

A) Коэффициент наклона показывает изменение зависимой переменной при единичном изменении независимой переменной. +

B) Коэффициент наклона показывает процентное изменение зависимой переменной при единичном изменении независимой переменной.

C) Коэффициент наклона показывает силу и статистическую значимость связи между зависимой и независимой переменными.

D) Коэффициент наклона показывает величину ошибки модели линейной регрессии.

E) Коэффициент наклона не имеет интерпретации в контексте линейной регрессии.

6. Определить, насколько хорошо модель линейной регрессии соответствует данным можно с помощью:

A) Сравнения коэффициента детерминации (R^2) с единицей. +

B) Сравнения среднеквадратической ошибки (MSE) с нулем.

C) Сравнения средней абсолютной ошибки (MAE) с нулем.

D) Сравнения коэффициента корреляции Пирсона с нулем.

E) Сравнения коэффициента наклона (slope) с нулем.

**Методичка: ML1.2 multivariate**

7. Основная характеристика модели множественной регрессии?

A) Она позволяет анализировать взаимосвязь между одной зависимой переменной и несколькими независимыми переменными. +

B) Она используется только для категоризации данных и прогнозирования качественных переменных.

C) Она предполагает линейные зависимости между всеми независимыми переменными.

D) Она не требует выполнения каких-либо предположений о распределении остатков.

E) Она идеально подходит для решения всех видов задач регрессионного анализа.

8. Каким образом можно обучить модель множественной регрессии в Python с использованием библиотеки scikit-learn?

A) Разделить данные на обучающую и тестовую выборки, создать экземпляр модели LinearRegression, обучить модель на обучающих данных и оценить ее производительность на тестовых данных. +

B) Использовать функцию fit() на экземпляре модели MultipleRegression из модуля sklearn.linear\_model, передавая в нее обучающие данные и целевую переменную.

C) Использовать функцию train() на экземпляре модели LinearRegression из модуля sklearn, передавая в нее обучающие данные и целевую переменную.

D) Использовать функцию fit() на экземпляре модели LogisticRegression из модуля sklearn.linear\_model, передавая в нее обучающие данные и целевую переменную.

E) Использовать функцию train() на экземпляре модели MultipleRegression из модуля sklearn.linear\_model, передавая в нее обучающие данные и целевую переменную.

**Методичка: ML1.3 real regression**

9. Каким образом можно улучшить производительность модели регрессии в задаче реальных чисел (real regression)?

A) Использовать более сложную модель с большим числом параметров.

B) Применить методы регуляризации, такие как L1 или L2 регуляризация, для контроля переобучения модели. +

C) Провести отбор признаков и оставить только наиболее информативные переменные.

D) Увеличить размер обучающего набора данных.

E) Применить методы аугментации данных для расширения обучающего набора данных.

10. Какая метрика обычно используется для оценки качества модели регрессии?

A) Precision

B) Recall

C) F1-score

D) R-squared +

E) Accuracy

**Методичка: ML2.1 logistic regression**

11. Каким образом происходит интерпретация коэффициентов в логистической регрессии?

A) Чем выше коэффициент, тем сильнее влияние признака на классификацию

B) Чем выше коэффициент, тем слабее влияние признака на классификацию

C) Знак коэффициента указывает на положительное или отрицательное влияние признака +

D) Коэффициенты не имеют интерпретации в логистической регрессии

E) Коэффициенты определяют порядок важности признаков

12. Какие значения может принимать выходная переменная в логистической регрессии?

A) Действительные числа

B) Целые числа

C) Бинарные значения (0 или 1) +

D) Категориальные значения

E) Множество действительных чисел

**Методичка: ML2.2 real classification**

13. Какой алгоритм классификации чаще всего используется для задачи "к ближайших соседей"?

A) K-means

B) Naive Bayes

C) Decision Tree (дерево решений)

D) K-nearest neighbors (KNN) +

E) Support Vector Machine (SVM)

14. Как работает алгоритм "случайного леса" в задаче классификации реальных данных?

A) Объединяет несколько деревьев решений в одну модель +

B) Определяет оптимальные разделяющие поверхности между классами

C) Использует вероятности классов для принятия решений

D) Применяет метод градиентного спуска для минимизации ошибки

E) Использует метод опорных векторов для классификации