Лабораторные работы

Лаб 1. Построение линейной регрессии. Вычисление коэффициента корреляции.

1. Построить линейную регрессию.
2. Вычислить коэффициент корреляции.

Вопросы:

1. Вывести формулу для коэффициентов уравнения регрессии.
2. Объяснить, почему в формуле оценки коэффициента корреляции в знаменателе (n – 1) (Подсказка: несмещенная оценка).

Лаб 2. Проверка гипотез.

1. Найти доверительный интервал для математического ожидания и дисперсии.
2. Проверить гипотезу о равенстве математических ожиданий двух генеральных совокупностей:

а) для известных дисперсий; б) для неизвестных дисперсий.

Вопросы:

1. Пояснить, как строится доверительный интервал.
2. Как проверяется гипотеза (в общем виде).
3. Привести примеры распределений, которые применяются для проверки гипотез.

Лаб 3. Описательная статистика для изображений.

1. Взять 2 произвольные .JPG изображения. Конвертировать изображения в полутоновые (grayscale). Построить гистограммы изображений. Найти выборочное среднее, выборочное среднеквадратическое, моду и медиану для гистограмм. Найти коэффициент корреляции для 1) гистограмм; 2) изображений.

*Указание*. В гистограмме использовать группировку интенсивностей, то есть разбить интервал интенсивностей [0, 255] на равные подинтервалы, напр длины 10 [0, 9), [10, 19) …, [240, 249), [250, 255) – последний интервал неполной длины, он учитывает только 6 интенсивностей.

1. Критерием Пирсона проверить гипотезы о соответствии выборочных функций распределений нормальному закону.

Вопросы:

1. Пояснить, что такое мода, медиана, выборочное среднее, выборочная дисперсия, выборочный коэффициент корреляции.
2. Как проверяется гипотеза (в общем виде). Проверка гипотезы о нормальности распределения (по Пирсону или Колмогорову-Смирнову).

Лаб 4. Кластерный анализ 1.

1. Разделить выборку, в которой на элементах измерено не менее трех параметров, методом k-средних (методом k-медианы).
2. Методом k-средних создать кластеры текстов. Каждый текст выбран из научной литературы и состоит из 3-х страниц. Создать словари терминов для каждой науки и найти количество терминов в тексте из каждого словаря.

Указания. Например, для 3-х научных направлений создать 3 словаря терминов (3 тезауруса). В каждом направлении выбрать не менее 5 текстов. Вектор признаков каждого текста будет иметь 3 компоненты, такой вектор строится для каждого текста. Из текста выделить термины 1-го словаря и записать в 1-ю компоненту вектора признаков количество таких терминов в этом тексте. То же выполнить для 2-го и 3-го словарей.

Если количество текстов равно n, то в результате анализа выборки будет построено n векторов (n>=15). Выполнить для них кластеризацию методом k-средний, здесь k = 3.

Лаб 5. Применение сети Кохонена для кластеризации.

1. Выбрать данные для задачи кластеризации. Минимум 200 элементов в выборке, не менее 5 признаков для элементов.
2. Построить нейронную сеть со слоем Кохонена (одномерным, двумерным или трехмерным).
3. Из выборки выбрать репрезентативное множество и провести обучение сети на этом множестве.
4. Выполнить кластеризацию элементов выборки. Пояснить полученный результат.

Вопросы:

1. Пояснить, как выполняется обучение сети.
2. Объяснить полученные результаты кластеризации.