Міністерство освіти і науки України

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «КИЄВО-МОГИЛЯНСЬКА АКАДЕМІЯ»

Факультет інформатики

Изображение выглядит как зеркало

Автоматически созданное описание

**Протокол до лабораторної роботи №1**

**З дисципліни „Математичні методи машинного навчання ”**

Виконав

студент 4 курсу

факультету інформатики

Картавий М. О.

Опис лабораторної роботи

1. Сформувати тестову вибірку зображень з вихідного пакету;
2. Для зеленого каналу кольору тестових зображень обчислити наступні характеристики розподілу значень яскравості пікселів:
   1. Математичне очікування та дисперсію;
   2. Медіану та інтерквартильний розмах;
   3. Коефіцієнти асиметрії та ексцесу (нормалізований);
   4. Гістограму значень яскравості пікселів (нормалізовану);
3. Провести обробку отриманих гістограм:
   1. Провести апроксимацію гістограм з використанням імовірнісних розподілів:
      1. Нормального (гаусового) розподілу;
      2. Розподілу Лапласа;
      3. Розподілу Стьюдента;
      4. Бета розподілу;
   2. Для кожної гістограми визначити найкращий тип апроксимуючого розподілу за критерієм мінімізації середньо квадратичного відхилення;
   3. Побудувати розподіл отриманих апроксимацій за видом апроксимуючого розподілу;
4. Побудувати багатовимірні гаусові моделі:
   1. Сформувати вектори параметрів розподілу значень яскравості пікселів тестових зображень;
      1. Математичне очікування;
      2. Математичне очікування та дисперсія;
      3. Математичне очікування, дисперсія та коефіцієнт асиметрії;
      4. Математичне очікування, дисперсія, коефіцієнти асиметрії та ексцесу;
   2. Визначити параметри багатовимірних гаусових моделей для кожної групи векторів, відобразити їх графічно та у вигляді таблиць.
5. Підготувати звіт за отриманими результатами лабораторної роботи.

Виконання лабораторної роботи

Робота була виконана на мові Python в форматі Jupyter Notebook та за допомогою середовища Jupyter.

1. Для формування вибірки було використано модуль random відповідним діапазоном значень.
2. Для обчислення загальновибіркових характеристик яскравості пікселів зеленого кольору було сформовано масив, отриманий шляхом забору початкового зображення зеленого каналу.
3. Числові характеристики даного розподілу мають наступний вигляд:

|  |  |
| --- | --- |
| Математичне очікування | 104.7 |
| Дисперсія | 4123 |
| Медіана | 98 |
| Інтерквартильний розмах | 92 |
| Коефіцієнт ексцесу | 0.6 |
| Коефіцієнт асиметрії | 0,33 |

1. Оскільки, мат. очікування більше за медіану, правий хвіст розподілу більший за лівий, що підтверджується додатнім коефіцієнтом асиметрії.
2. Розподіл є між нормальним та гострим, коефіцієнту ексцесу не великий і не малий. Маємо велику дисперсію.

Обробка та апроксимація отриманих гістограм

Для вибірки з 250 зображень розподіл отриманих апроксимацій за видом апроксимуючого розподілу (за критерієм мінімізації середньоквадратичного відхилення):

Изображение выглядит как снимок экрана

Автоматически созданное описание

Бачимо, що отриманий розподіл має два виражених головних елемента – бета розподіл та лапласа.

Багатовимірні гаусові моделі

Сформувавши вектори параметрів розподілу значень яскравості пікселів для тестових зображень, ми сформували наступні гаусові моделі:

Математичне очікування

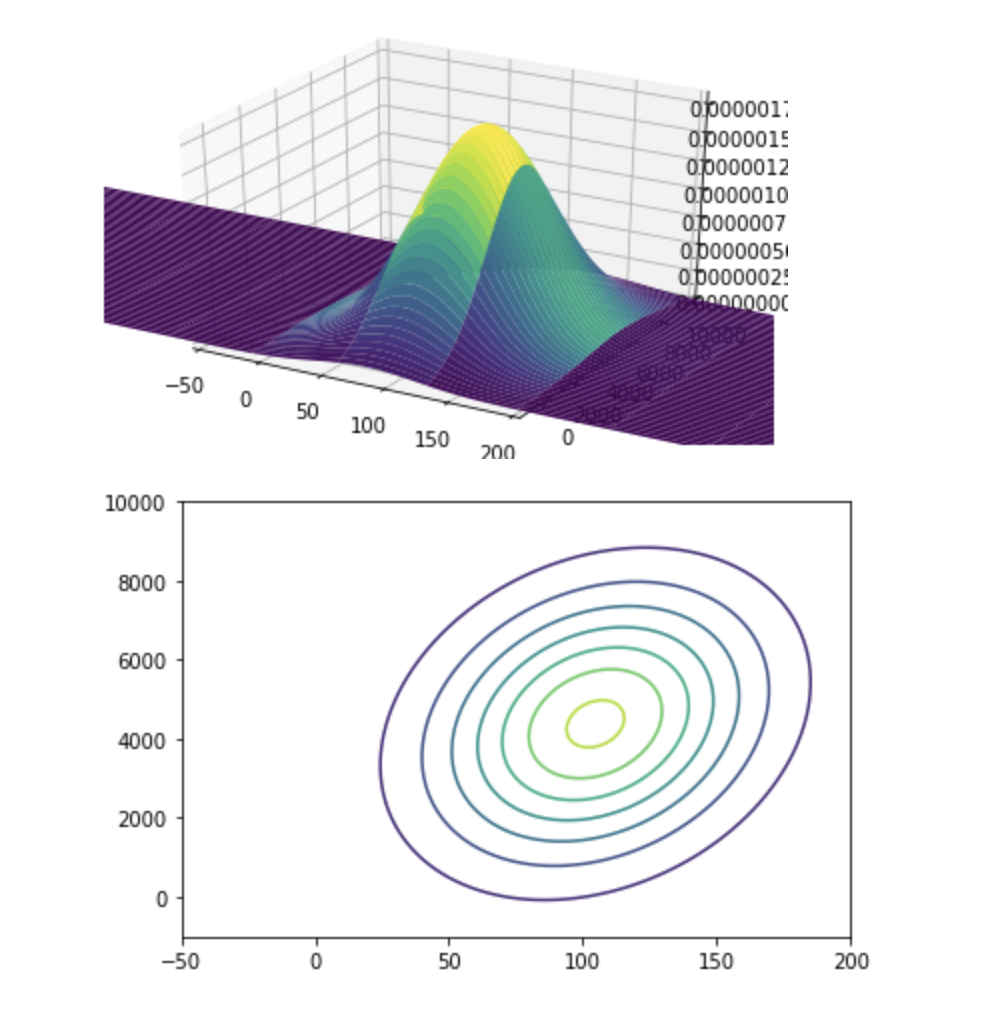
|  |  |
| --- | --- |
| Вектор середнього | 104.6 |
| Матриця коваріації | 1627 |

Изображение выглядит как снимок экрана

Автоматически созданное описание

Математичне очікування та дисперсія

|  |  |
| --- | --- |
| Вектор середнього | 104.7 ; 4383 |
| Матриця коваріації | [[ 1627 21343]  [21343 5013292]] |



Математичне очікування, дисперсія та коефіцієнт асиметрії

|  |  |
| --- | --- |
| Вектор середнього | 104.7 ; 4383; 0.41 |
| Матриця коваріації | [[ 1627 21343 -37]  [ 21343 5013292 -481 ]  [ -37 -481 -1.26]] |

Математичне очікування, дисперсія, коефіцієнти асиметрії та ексцесу

|  |  |
| --- | --- |
| Вектор середнього | 104 4383 0.416 0.688 |
| Матриця коваріації | [[ 1627 21343 -37 -87]  [ 21343 5013292 -481 -4687]  [ -37.87 -481.18 1.26 4.9]  [ -87.33 -4687.32 4.9 42 ]] |

Висновки

1. З отриманих даних MIRFlickr я відібрав випадковим чином 250 картинок. Відібрав зелений канал та провів розрахунок наступних характеристик: математичного сподівання, дисперсії, медіани, інтерквартального розмаху, коефіцієнтів асиметрії та ексцесу. Потім візуалізував гістограми значень яскравості пікселів та апроксимував за допомогою одного із зазначених розподілів: Нормального, Лапласа, Стьюдента, Бета за критерієм мінімізації середньо квадратичного відхилення. Потім побудував багатовимірні гаусові моделі з різними вхідними векторами. На основі розрахованих даних можна сказати, що у тестовій вибірці було зосереджено лапласа та бета розподіли отриманих апроксимацій за видом апроксимуючого розподілу. А також вибірка має розподіл між нормальним та гострим, коефіцієнту ексцесу не великий і не малий. Маємо велику дисперсію.