Міністерство освіти і науки України

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «КИЄВО-МОГИЛЯНСЬКА АКАДЕМІЯ»

Факультет інформатики



**Протокол до лабораторної роботи №1**

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

Виконав

студент 4 курсу

факультету інформатики

Крошин О. А.

1. Завдання 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.

2. Порядок виконання роботи та отримані результати

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

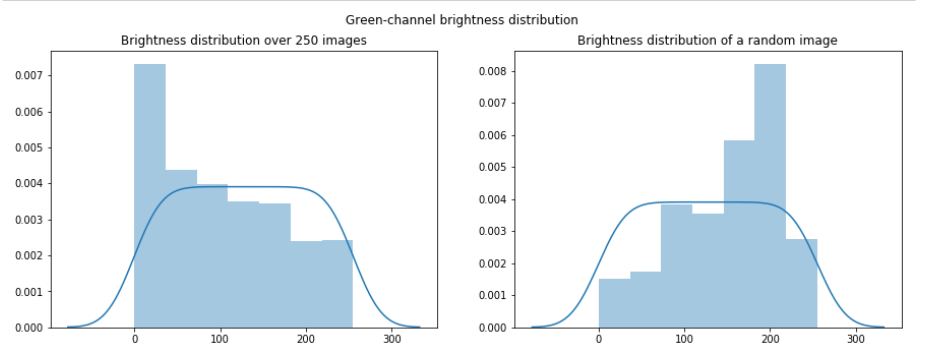
1. Для формування вибірки було використано модуль random теки numpy з відповідним зафіксованим сідом (початковим кроком).
2. Для обчислення загальновибіркових характеристик яскравості пікселів зеленого кольору було сформовано масив довжини 256 з відповідними частотами кожного значення яскравості пікселів зеленого каналу.

Рис.1. Нормалізована гістограма яскравості пікселів зеленого каналу. ліворуч – для побудованої вибірки, праворуч – для випадкової картинки з вибірки

Числові характеристики даного розподілу (ліворуч) мають наступний вигляд:

|  |  |
| --- | --- |
| Математичне очікування | 100.92 |
| Дисперсія | 5702.30 |
| Медіана | 90.0 |
| Інтерквартильний розмах | 127.0 |
| Коефіцієнти асиметрії | 0.3841 |
| Коефіцієнт ексцесу | -0.9989 |

Бачимо, що мат. Очікування більше за медіану, правий хвіст розподілу більший за лівий, що підтверджується додатнім коефіцієнтом асиметрії.

Пік розподілу є відносно гладким (platykurtic), що не суперечить від’ємному коефіцієнту ексцесу. Маємо відносно велику дисперсію, що не є чимось дивним.

1. Обробка та апроксимація отриманих гістограм

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

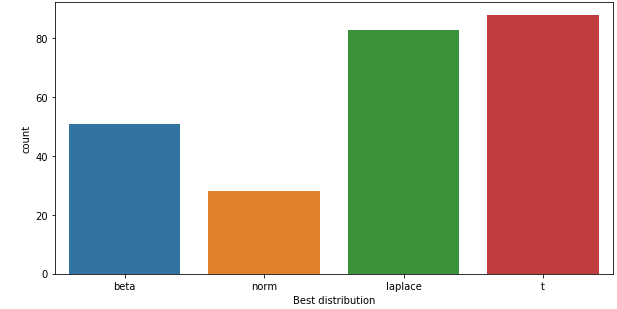


Рис. 2. Розподіл отриманих апроксимацій за видом оптимального апроксимуючого розподілу

Бачимо, що отриманий розподіл не має вираженого головного елемента. Так, розподіли Лапласа та Т-розподіл Стьюдента є оптимальними частіше за бета-розподіл та нормальний розподіл, але їхні частоти є приблизно рівними.

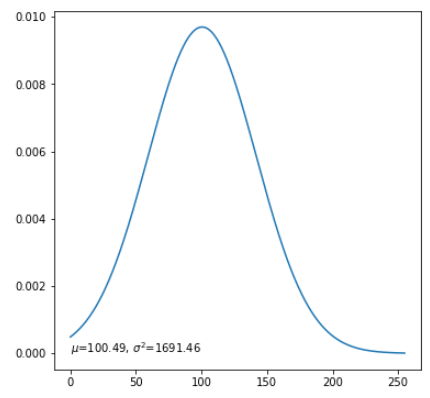
1. Багатовимірні гаусові моделі

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

* + - Математичне очікування

|  |  |
| --- | --- |
| Вектор середнього | 100.4983 |
| Матриця коваріації | 1691.4640 |

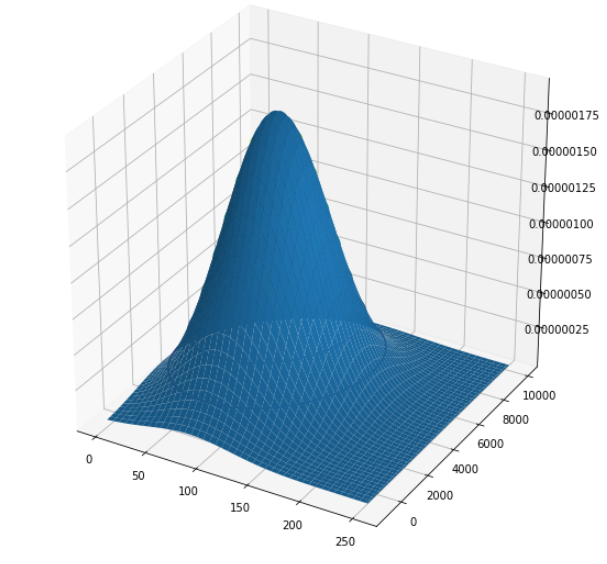
Табл.1.Параметри багатовимірної Гаусової моделі побудованій на мат. очікуванні



* + - Математичне очікування та дисперсія

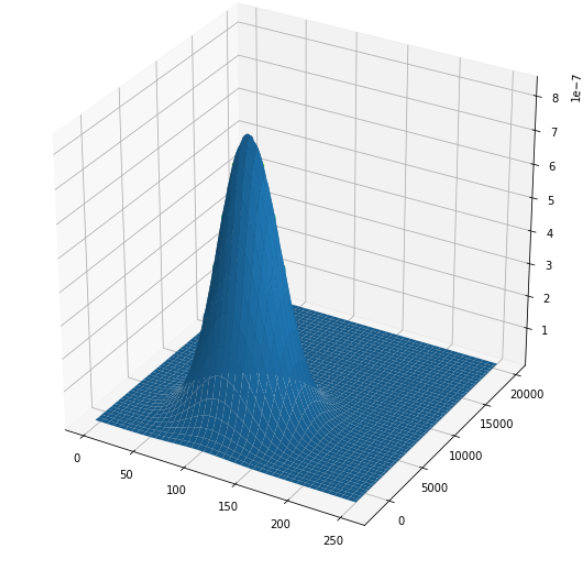
|  |  |
| --- | --- |
| Вектор середнього | [ 100.4928 4000.8261] |
| Матриця коваріації | [[ 1691.464 15428.9177]  [ 15428.9177 4113532.6315]] |

Табл.2.Параметри багатовимірної Гаусової моделі побудованій на мат. очікуванні та дисперсії



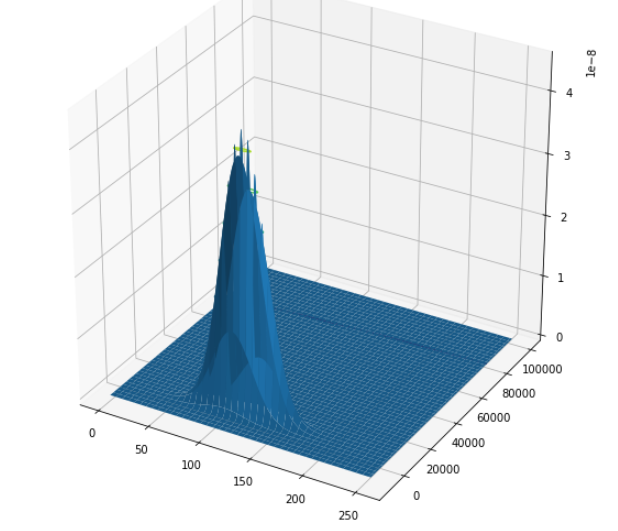
* + - Математичне очікування, дисперсія та коефіцієнт асиметрії

|  |  |
| --- | --- |
| Вектор середнього | [ 100.4928 4000.8261 0.4765] |
| Матриця коваріації | [[ 1691.464 15428.9177 -42.434]  [ 15428.9177 4113532.6315 -439.662]  [-42.434 -439.6632 1.8177]] |



* + - Математичне очікування, дисперсія, коефіцієнти асиметрії та ексцесу

|  |  |
| --- | --- |
| Вектор середнього | [ 100.4928 4000.8261 0.4765 1.4513] |
| Матриця коваріації | [[ 1691.464 15428.9177 -42.434 -124.9112]  [ 15428.9177 4113532.6315 -439.662 -6406.975]  [-42.434 -439.6632 1.8177 12.7983]  [-124.9112 -6406.975 12.7983 191.5563]] |



3. Висновки

В ході лабораторної роботи ми проаналізовали розподіл яскравості пікселів зеленого каналу згенерованої вибірки запропонованого нам пакету картинок. Сформувавши вибірку та обчисливши числові характеристики розподілів яскравості відповідних картинок, ми побачили, що розподіл їхніх значень істотно відрізняється поміж картинками. Мінімізація похибки за допомогою критерію середньоквадратичного відхилення продемонструвала, які розподіли краще апроксимують певні отримані гістограми. Побудова апроксимуючого розподілу продемонструвала, що не можна сказати, що якісь з запропонованих для апроксимації розподілів (Лапласа, Нормальний, Бета, Т-розподіл Стьюдента) апроксимує нормалізовані гістограми яскравості краще за інших. Також по векторах параметрів розподілів яскравості зеленого каналу кожної картинки ми побудували багатовимірні гаусові моделі та визначили їхні вектори середнього та матриці коваріації.