

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «КИЄВО-МОГИЛЯНСЬКА АКАДЕМІЯ»

Факультет інформатики

Кафедра математики

ЛАБОРАТОРНА РОБОТА №1

з курсу «Математичні методи машинного навчання»

Варіант 2

Виконала:

студентка 3 р.н.

БП «Прикладна математика»

Гак Софія Володимирівна

Київ-2020

Мета:

формування вибірки із 25-тисячної генеральної сукупності зображень, що в результаті складає всього 250 зображень (1). Обчислення необхідних характеристик (2) для подальшої апроксимації розподілу значень яскравості пікселів зеленого каналу зображень вибіркового пакету відомими ймовірнісними розподілами (нормальним, розподілом Лапласа, Стюдента, бета-розподілом) (3), а також багатовимірним аналогом нормального розподілу (4).

Частка виконаної роботи:

Повністю виконана: п.1-4, а також одне із додаткових завдань з книги Murphy.

1) Сформувати тестову вибірку зображень з вихідного пакету;

Для псевдовипадкового формування вибірки зображень застосовано звичайну функцію `random()`, оскільки в її основі лежить генератор Мерсена (згідно з документацією). Стартове значення – 2 (за номером студента в списку групи). Формування доволі тривіальне: спершу генерується 250 чисел із 25000, потім з архіву один раз вилучається папка `mirflickr` на Google Диск із зображеннями із номерами, відповідним цим 250-м числам, які у свою чергу конвертуються в `numpy`-масиви і в подальшому зберігаються в іншому масиві із назвою `imgs`.

2) Для зеленого каналу кольору тестових зображень обчислити наступні характеристики розподілу значень яскравості пікселів:

Для розрахунку характеристик було використано їх реалізації в бібліотеці `numpy`.

а) Математичне очікування та дисперсію;

Математичне очікування: 101.77400161065259 (тобто майже середина допустимого діапазону значень яскравості пікселів: 0-255).

Дисперсія: 5545.282938782499

б) Медіану та інтерквартильний розмах;

Медіана: 96.0

Інтерквартильний розмах: 126.0

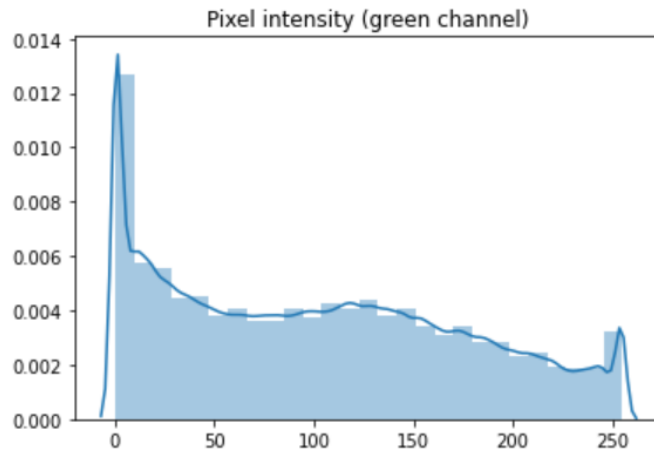
Інтерквартильний розмах був обчислений як різниця між 25-м і 75-м перцентилями (або 1-м і 3-м квантилями), що описує розмах 50% значень яскравості пікселів, що знаходяться навколо медіани.

в) Коефіцієнти асиметрії та ексцесу (нормалізований);

Коефіцієнт асиметрії: 0.3212877038756634 – вказує на наявність у гістограми розподілу правого довгого “хвосту” порівняно із нормальним.

Нормалізований коефіцієнт ексцесу: -4.0173482995522924 – вказує на пологість вершини гістограми розподілу порівняно із нормальним.

г) Гістограму значень яскравості пікселів (нормалізовану);



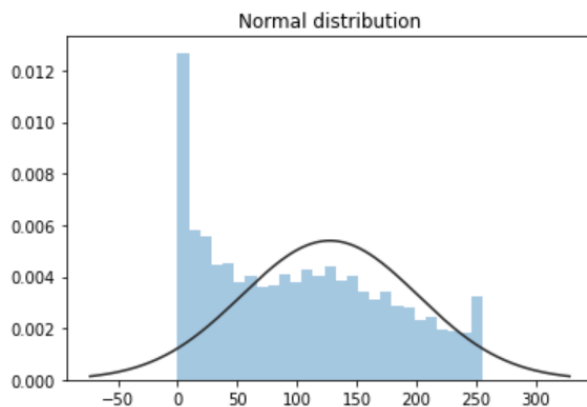
Згідно із гістограмою в зеленому каналі тестових зображеннях найбільше значень яскравості пікселів, близьких до 0, тобто зовсім темних відтінків зеленого, а то й чорного кольору.

3) Провести обробку отриманих гістограм:

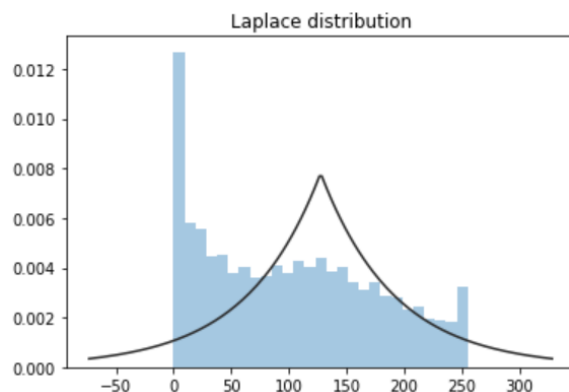
а) Провести апроксимацію гістограм з використанням імовірнісних розподілів:

Апроксимації були відтворені лише візуально (кількість інтервалів була розрахована за) для розподілу значень яскравості пікселів зеленого каналу ВСІХ зображень.

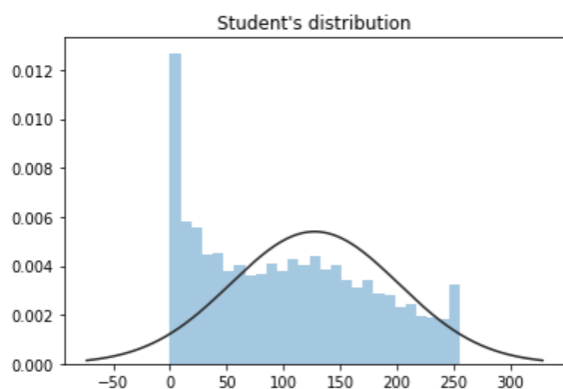
i) Нормального (гаусового) розподілу;



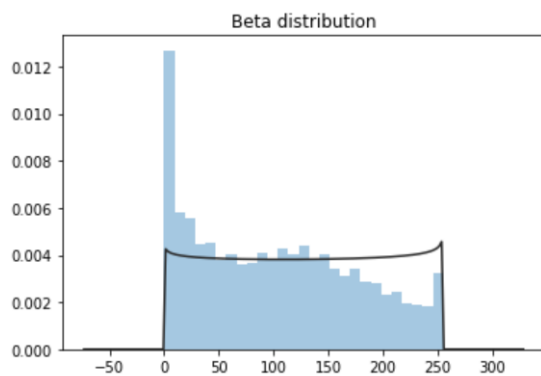
ii) Розподілу Лапласа;



iii) Розподілу Стюдента;



iv) Бета розподілу;



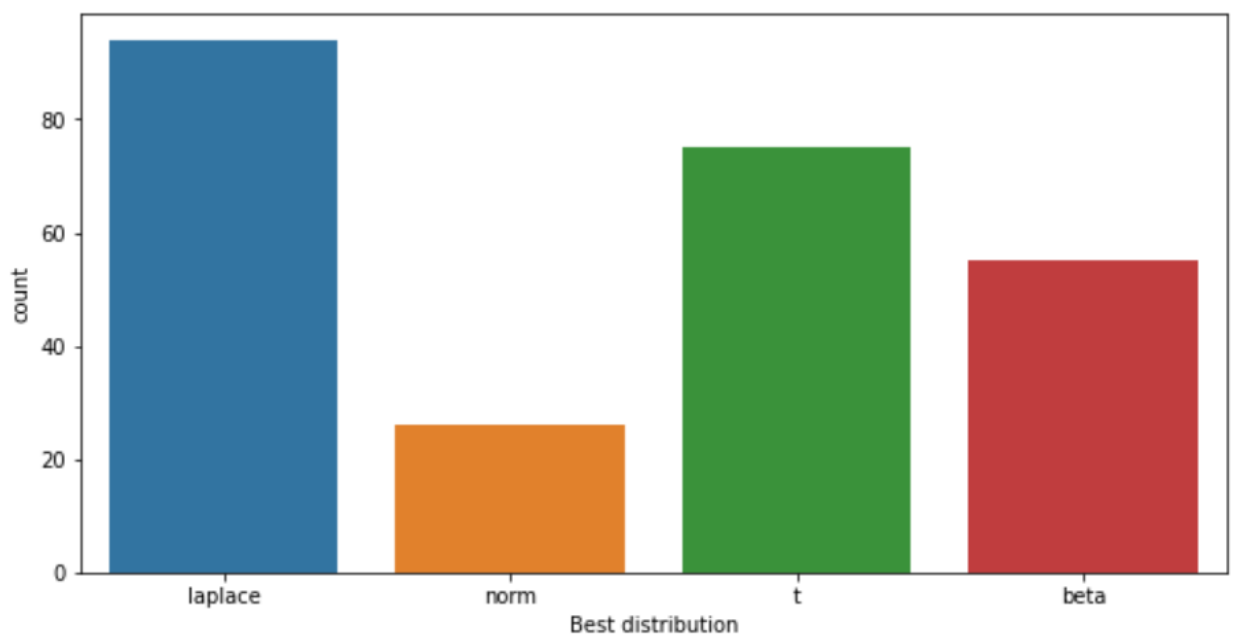
Подальші пункти, окрім, звичайно, підготовки звіту, не були реалізовані. Натомість були виконані деякі додаткові завдання (розписано у файлі із лабораторною роботою).

- b)** Для кожної гістограми визначити найкращий тип апроксимуючого розподілу за критерієм мінімізації середньо квадратичного відхилення;

В таблиці нижче наведено найкращий тип апроксимуючого розподілу для 5 перших і 5 останніх зображень тестової вибірки, а також показник середньо квадратичного відхилення.

	Best distribution	MSE
Image		
1	laplace	1.602494e+06
2	norm	7.294128e+05
3	t	9.334838e+05
4	laplace	1.124676e+06
5	norm	5.357496e+05
...
246	t	1.267409e+06
247	t	6.403343e+05
248	laplace	5.834728e+05
249	laplace	9.534430e+06
250	beta	9.161754e+05

- c)** Побудувати розподіл отриманих апроксимацій за видом апроксимуючого розподілу;



За даною гістограмою можна зробити висновок, що в більшості випадків, найкращим типом апроксимуючого розподілу є лапласовий.

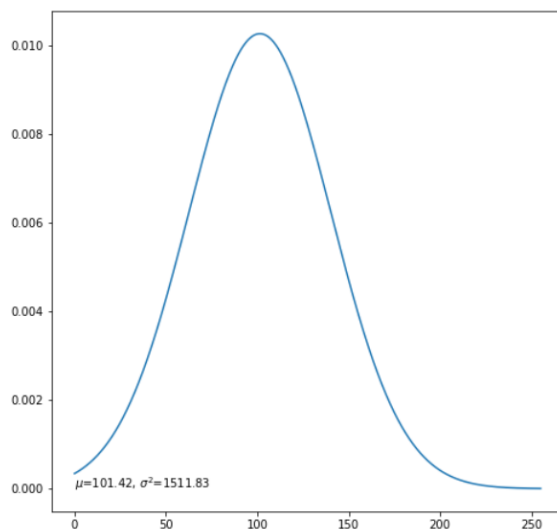
4) Побудувати багатовимірні гаусові моделі:

- a) Сформувати вектори параметрів розподілу значень яскравості пікселів тестових зображень;
- b) Визначити параметри багатовимірних гаусових моделей для кожної групи векторів, відобразити їх графічно та у вигляді таблиць.

Всі вектори розраховуються в окремій функції в залежності включення в її аргументи ту чи іншу характеристику булевим значенням (True/False).

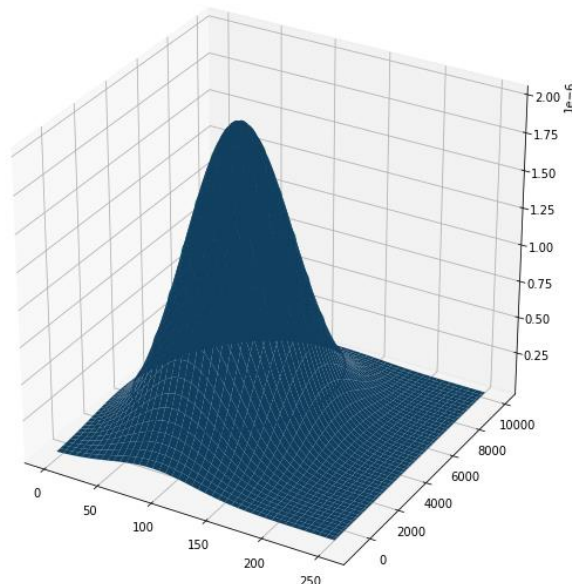
Параметри моделей обраховуються тією самою функцією. Для кожної групи векторів, обрахованих в пункті a) отримали такі параметри і їх графічні подання:

i) Математичне очікування;



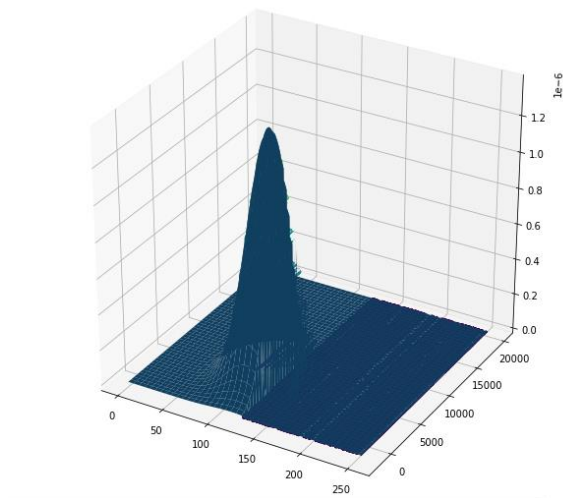
ii) Математичне очікування та дисперсія;

```
Mean vector:  
[ 101.41725486 4029.99573068],  
Covariance matrix:  
[[1.51183214e+03 1.06073339e+04]  
 [1.06073339e+04 4.20431174e+06]]
```



iii) Математичне очікування, дисперсія та коефіцієнт асиметрії;

```
Mean vector:  
[1.01417255e+02 4.02999573e+03 4.32624092e-01],  
Covariance matrix:  
[[ 1.51183214e+03 1.06073339e+04 -3.17161316e+01]  
 [ 1.06073339e+04 4.20431174e+06 -2.70671303e+02]  
 [-3.17161316e+01 -2.70671303e+02 9.40015945e-01]]
```



iv) Математичне очікування, дисперсія, коефіцієнти асиметрії та ексцесу;

```
Mean vector:  
[1.01417255e+02 4.02999573e+03 4.32624092e-01 4.57618704e-01],  
Covariance matrix:  
[[ 1.51183214e+03 1.06073339e+04 -3.17161316e+01 -5.46383876e+01]  
 [ 1.06073339e+04 4.20431174e+06 -2.70671303e+02 -3.27512736e+03]  
 [-3.17161316e+01 -2.70671303e+02 9.40015945e-01 2.20800721e+00]  
 [-5.46383876e+01 -3.27512736e+03 2.20800721e+00 1.49354792e+01]]
```

