

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:

Пункти 1, 2, частина 3, а також одне із додаткових завдань з книги Murphy.

Самоаналіз:

Загалом робота вимагає кращого, зокрема, це стосується відсутності виконання п. 4, а також недосконалої реалізації п. 3.

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

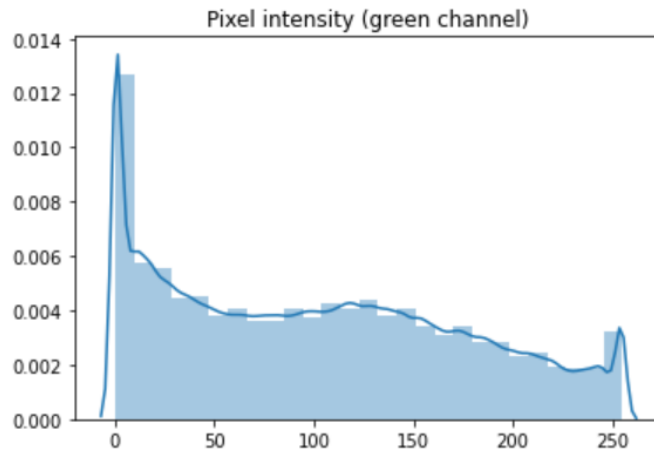
Інтерквартильний розмах був обчислений як різниця між 25-м і 75-м перцентилями (або 1-м і 3-м кuartилями), що описує розмах 50% значень яскравості пікселів, що знаходяться навколо медіани.

с) 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.

д) Гістограму значень яскравості пікселів (нормалізовану);



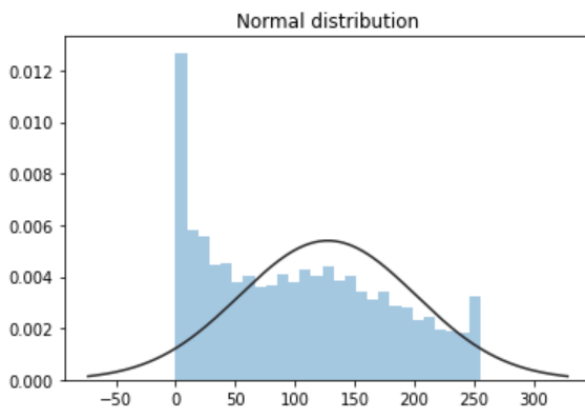
Згідно із гістограмою в зеленому каналі тестових зображеннях найбільше значень яскравості пікселів, близьких до 0, тобто зовсім темних відтінків зеленого, а то й чорного кольору.

3) Провести обробку отриманих гістограм:

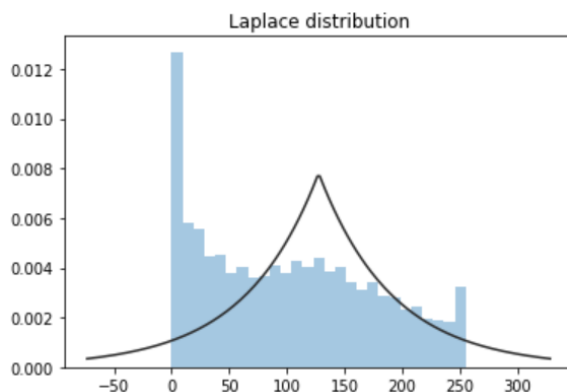
а) Провести апроксимацію гістограм з використанням імовірнісних розподілів:

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

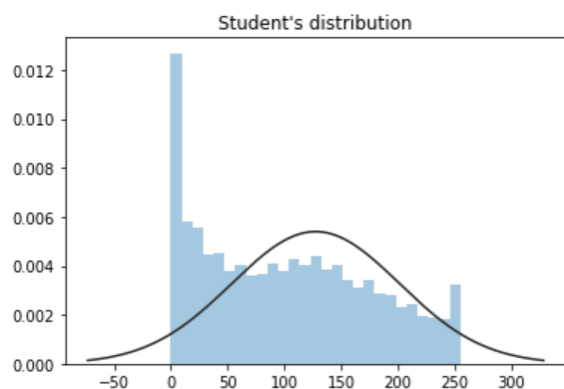
i) Нормального (гаусового) розподілу;



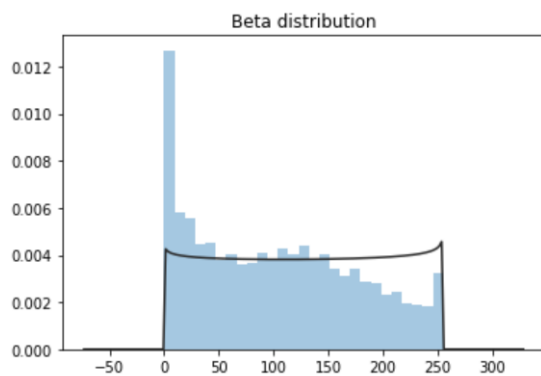
ii) Розподілу Лапласа;



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).