

Лабораторна робота №2

Андрій Пишко

Варіант 22 (за загальним списком)

1 Теоретична частина

1.1 Постановка завдання

Маємо три множини точок (дві з них генеруються згідно рівномірного розподілу, а інша - нормального). Згенерувати 100 точок, класифікувати кожну з них. Для визначення класу точки використати поняття відстані Махаланобіса

1.2 Ідея методу

Відстань Махаланобіса – міра відстані між векторами випадкових величин, що узагальнює поняття евклідової відстані.

Крок 1. Побудова моделі. Спочатку вимагається побудова моделі на основі навчальних даних. Це означає, що для кожного класу будуть зібрані зразки даних, і для цих зразків буде обчислена коваріаційна матриця.

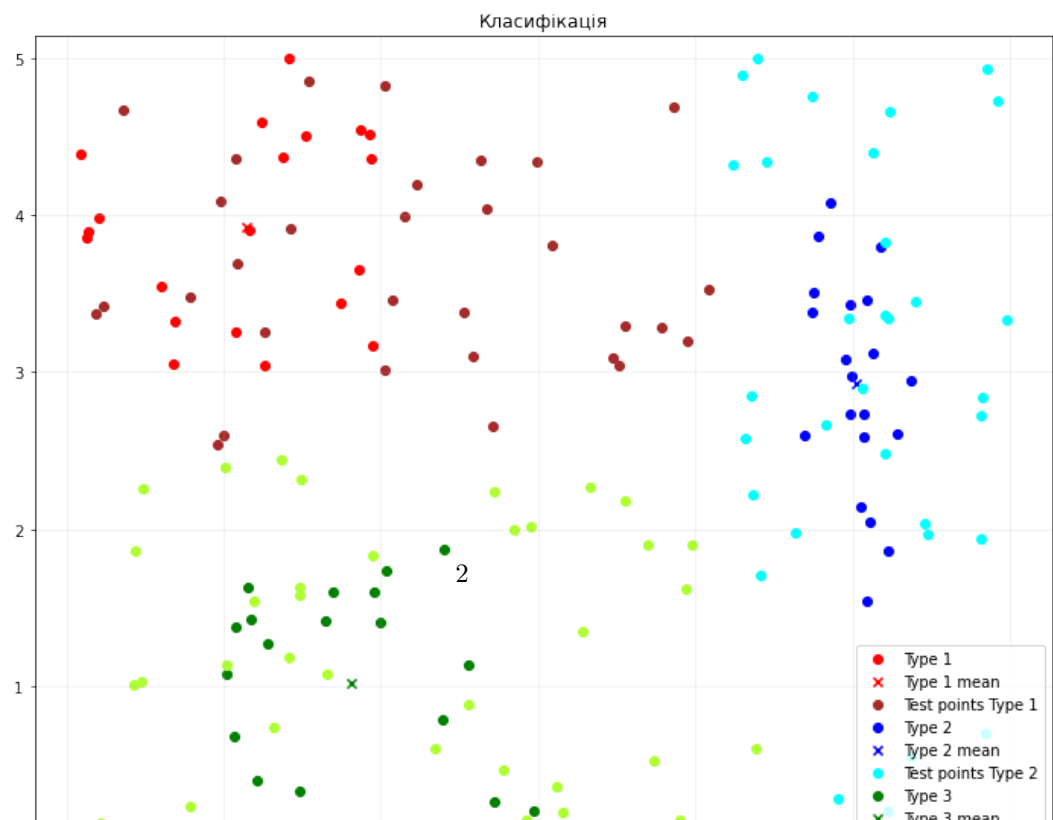
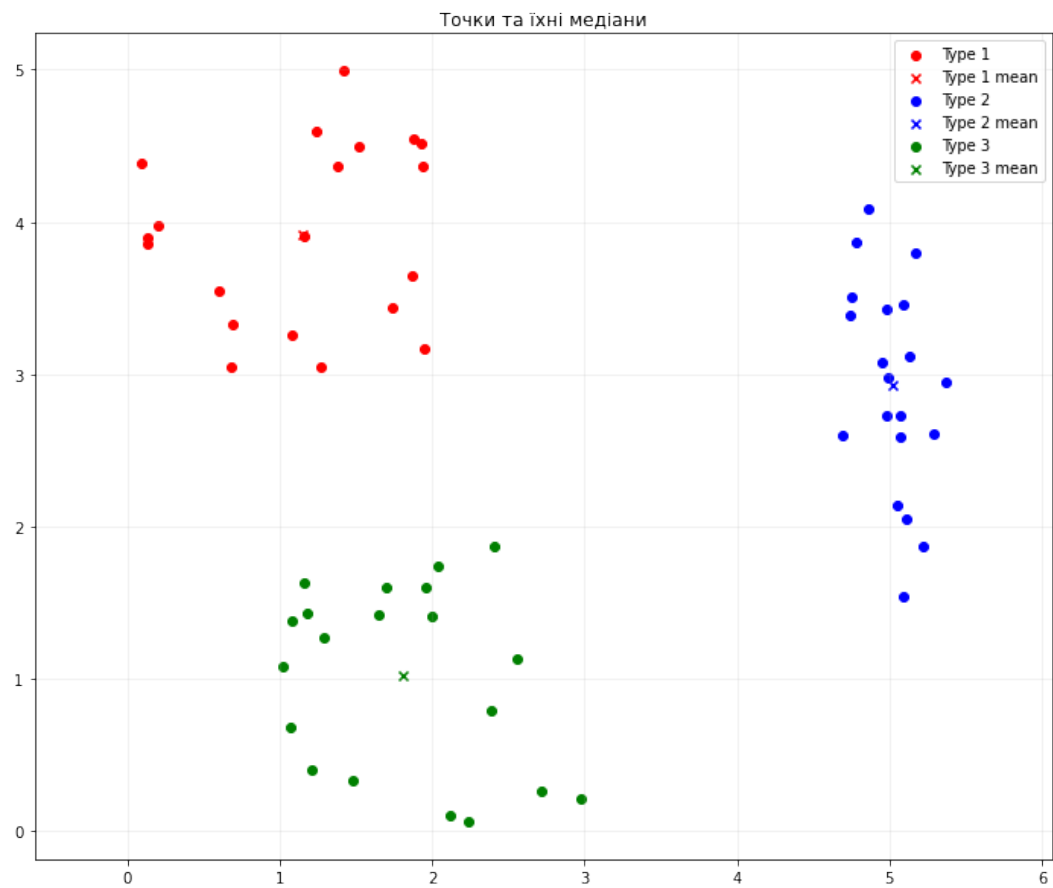
Крок 2. Обчислення відстані. Після побудови моделі можна обчислити відстань Махаланобіса для нових зразків даних. Для цього спочатку потрібно обчислити відстань між новою точкою і центром мас класу за допомогою формули:

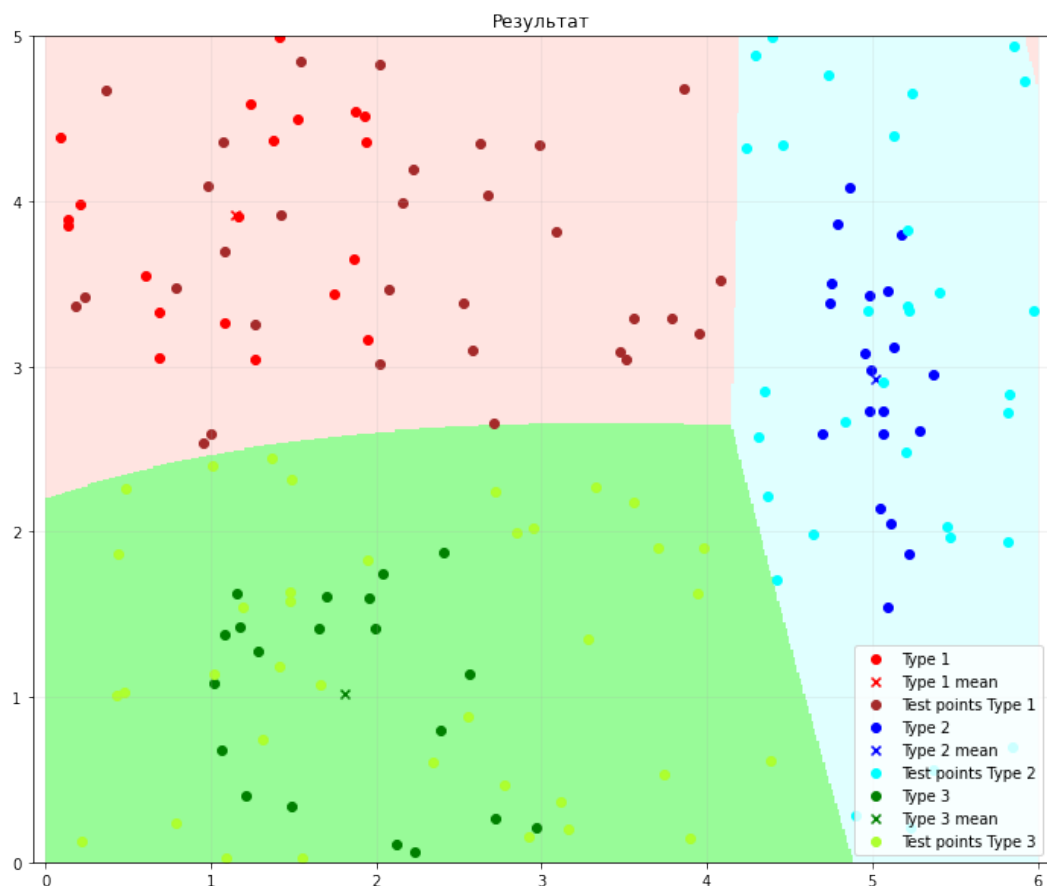
$$D^2 = (x - \mu)^T \Sigma^{-1} (x - \mu)$$

де D – відстань Махаланобіса, x – нова точка даних, μ – вектор середніх значень ознак для класу, Σ – коваріаційна матриця.

Крок 3. Поріг рішення. Після обчислення відстані можна встановити певний поріг, який визначає, до якого класу належить новий зразок даних. Якщо відстань Махаланобіса менше порогу, то зразок відноситься до даного класу, інакше – до іншого класу.

2 Результати





3 Висновок

Метод класифікації за відстанню Махаланобіса базується на мультівимірній статистиці. Він враховує кореляцію і дисперсію даних у великому числі розмірностей. Цей метод використовує матрицю коваріації, яка відображає залежність між різними ознаками даних. Він вимірює відстань між точками даних і центром кожного класу, використовуючи матрицю коваріації. Об'єкт призначається класу, до якого він має найменшу відстань від центру класу за Махаланобісовою відстанню.

Порівняю його з попереднім методом (з еліпсами Петуніна). По-перше, метод еліпсів Петуніна використовує геометричний підхід, побудовуючи еліпси, що охоплюють об'єкти кожного класу. Метод відстані Махаланобіса ґрунтується на статистичних властивостях даних, враховуючи кореляцію і дисперсію ознак. По-друге, метод еліпсів Петуніна може бути більш простим у використанні, але не враховує взаємозв'язки між ознаками. Метод відстані Махаланобіса ураховує кореляцію і дисперсію ознак, що може бути

корисним у випадках, коли дані мають складну структуру.