

## Computer Homework 3

จัดทำโดย

นายสุริยา เตชะลือ

รหัสนักศึกษา 600610790

เสนอ

ผศ.ดร. ศันสนีย์ เอื้อพันธุ์วิริยะกุล

รายงานนี้เป็นส่วนหนึ่งของรายวิชา 261456

(Introduction to Computational Intelligence)

## สารบัญ

สารบัญ	1
รูปแบบการทดลอง	2
การเตรียมการข้อมูล	2
การทำงานของระบบ	2
การทดลอง	4
สรุปผลการทดลอง	5

## รูปแบบการทดลอง

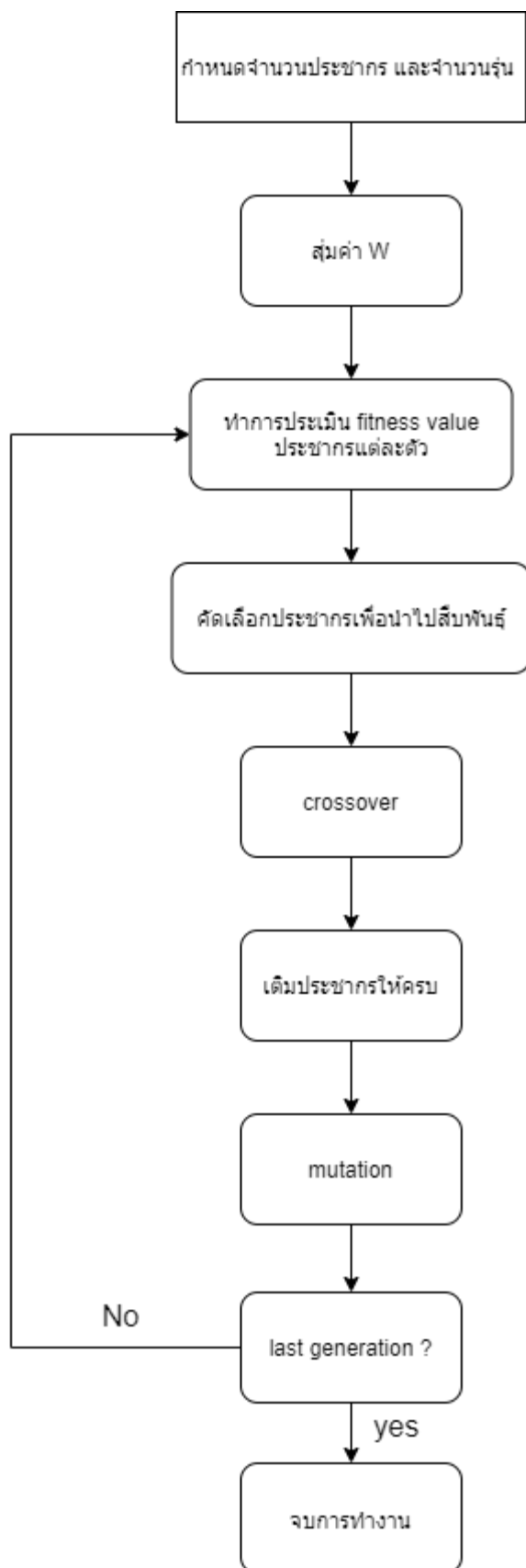
### 1. การเตรียมการข้อมูล

ในการทดลองนี้ได้ใช้ข้อมูล wdbc.data ซึ่งเป็นข้อมูล Wisconsin Diagnostic Breast Cancer จาก UCI Machine learning Repository) โดยมีข้อมูล sample ละ 32 ค่า ประกอบด้วย 1 id , 1 class และ 30 feature ซึ่งในการทดลองนี้ได้ทำการ classification โดยมี output คือ 1 class และมี input คือ 30 feature โดยทำการสร้าง Neural network ด้วยกระบวนการ Genetic Algorithms

ในการทดลองนี้ได้ทำการ นำ feature ทั้ง 30 ค่า ไปทำการ Max-min normalization ให้อยู่ในช่วง [0,1] และแปลงข้อมูลทั้ง 2 class โดยให้ M มีค่าเป็น 1 และ B มีค่าเป็น 0

### 2. การทำงานของระบบ

- กำหนดจำนวนของประชากรที่จะมีในแต่ละรุ่น จำนวน hidden layer ที่อยู่ในตัวประชากร และจำนวน generation ที่จะทำการทดลอง
- ทำการสุ่มค่า weight ในประชากรแต่ละคน โดยสร้างประชากรจำนวนเท่ากับจำนวนประชากรในแต่ละรุ่น
- ทำการประเมินแต่ละประชากรถึง fitness value โดยจะนำประชากรที่มี fitness value ที่ดีไปใช้ต่อไป
- ทำการคัดเลือกประชากรที่ดีที่สุดจากค่า fitness โดยใช้ Rank-based Selection ซึ่งเลือกประชากรที่ดีที่สุดจำนวนแรกๆ ไปสืบพันธุ์ต่อไป
- ทำการจับคู่ประชากรที่ได้เลือกไปในข้อที่ผ่านมา ทำการ Crossover เพื่อได้ประชากรรุ่นถัดไป
- ทำการเติมประชากรรุ่นถัดไปให้ครบตามจำนวนประชากรที่จะมีได้ในแต่ละรุ่น โดยทำการคัดลอกจากประชากรที่ดีอันดับต้นๆ ของรุ่นที่ผ่านมา
- ทำการ mutation ประชากรรุ่นถัดไปโดยวิธีการสุ่ม โดยใช้วิธีการ strong mutation ในการทำ mutation
- นำประชากรที่ได้ในข้อที่ผ่านมา กลับไปทำการหาค่า fitness และสืบพันธุ์ต่อไป จนกว่าจะถึง generation สุดท้าย
- นำประชากรที่ดีที่สุดใน generation สุดท้ายเป็นคำตอบ



แผนผังแสดงการทำงานของโปรแกรม

### 3. การทดลอง

กำหนดตัวแปรต้นของการทดลอง โดยมีการสุ่มค่า weight ในแต่ละประชากรในช่วง  $[-3,3]$  ซึ่งมีประชากร 100 ตัว ทำการทดลอง 1000 รุ่น และใช้ Mean square error ในการหา fitness value ดังนี้

ตัวแปร	ค่าที่ใช้
สุ่มค่า weight ในช่วง	$[-3,3]$
จำนวนประชากรในแต่ละรุ่น	100
จำนวนรุ่น	1000
Objective function	$f(x) = \frac{1}{\log(1 + \sum(y - d)^2)}$

กำหนดตัวแปรตามในการทดลองแต่ละครั้งมีการปรับเปลี่ยนจำนวน hidden layer และจำนวน node ทั้งหมด 4 โครงสร้าง ดังนี้

โครงสร้างที่	Neural name
1	30-10-1
2	30-20-1
3	30-20-30-1
4	30-30-30-1

## สรุปผลการทดลอง

การใช้ Rank-based Selection ที่เลือกประชากรที่ดีที่สุดใน 50 อันดับแรกมาออกลูก อาจจะได้ผลลัพธ์ที่มีค่าคำตอบกระจุกตัว ในถิ่นกำเนิด ซึ่งอาจจะไม่ได้คำตอบที่แท้จริง อีกทั้งทำให้โมเดลที่สร้างขึ้นมาไม่มีความยืดหยุ่น