LAB 2 Distance-based Classification

ชื่อ-สกุล..... รหัสนักศึกษา..... ตอนเรียน .....

จากไฟล์ Pima_diab.csv ตอบคำถาม ข้อ 1 – 10 ต่อไปนี้

1. อ่านไฟล์ข้อมูลดังกล่าว แสดงข้อมูล 20 แถวแรก
2. จำนวน samples ของผู้ป่วยที่เป็นเบาหวาน (tested_positive) ..... และจำนวน samples ของผู้ป่วยที่ไม่ได้เป็นเบาหวาน (tested_negative) .....
3. สร้างฟังก์ชั่น (function) เพื่อ clean ชุดข้อมูลนี้ ดังต่อไปนี้

   3.1 ลบคอลัมน์ 'unnamed:0'

   3.2 เติม missing value ด้วยค่าเฉลี่ย (mean) ในแต่ละคอลัมน์ประเภท numeric และเติมด้วย most frequent value สำหรับคอลัมน์ประเภท categorical

   3.3 แทนค่าในคอลัมน์ class โดยแทน 'tested_negative' ด้วย 0 และ 'tested_positive ด้วย 1

   โดยส่งค่ากลับเป็น dataframe ที่ถูก clean แล้ว

4. สร้างฟังก์ชั่นเพื่อเลือก features ที่มีค่า correlation กับ class สูงสุด 4 อันดับแรก พร้อมนำมาสร้าง dataframe ใหม่ ตั้งชื่อ newdf โดย feature ที่ถูกเลือกคือ .....
5. จากผลลัพธ์ dataframe ในข้อที่ 4 แบ่งข้อมูลสำหรับฝึกฝน (training data) ออกเป็น 70% และข้อมูลสำหรับทดสอบ (test data) 30% ให้แสดงรูปร่างมิติของข้อมูลฝึกฝน ..... และข้อมูลทดสอบ .....
6. สร้างโมเดล K-NN โดยกำหนดให้ k = 5 จากนั้นแสดงความแม่นยำบนชุดข้อมูลทดสอบ .....
7. ทำนาย label ของข้อมูลแถวที่ 3 บนชุดข้อมูลทดสอบ ลาเบลที่โมเดลทำนาย .....
8. แสดงรายการของ training sample 5 อันดับแรก ที่อยู่ใกล้กับข้อมูลทดสอบแถวที่ 3

   .....

9. ใช้ GridSeaerchCV เพื่อหาค่า k ที่เหมาะสม โดยเริ่มจาก k = 1 ถึง 20 บนชุดข้อมูลฝึกฝน ค่า k ที่ถูกเลือกโดย GridSearch คือ .....
10. แสดงความแม่นยำของโมเดลที่ใช้ค่า k ที่ได้รับจากข้อ 9 .....
11. บันทึกโมเดลข้อ 9 ในรูปของไฟล์ pickle ตั้งชื่อ knn.pkl

จากชุดข้อมูลภาพใบหน้า fetch_olivetti _faces ใน sk-learn ตอบคำถามข้อที่ 12 ถึง 16

12. แสดงรูปร่างมิติของชุดข้อมูลดังกล่าว ..... และจำนวน label .....
13. แบ่งชุดข้อมูลนี้ออกเป็น 80% สำหรับข้อมูลฝึกฝน และ 20% สำหรับข้อมูลทดสอบ โดยจำนวนข้อมูลฝึกฝนเท่ากับ ..... และจำนวนข้อมูลทดสอบ .....

LAB 2 Distance-based Classification

ชื่อ-สกุล..... รหัสนักศึกษา..... ตอนเรียน .....

14. สร้างโมเดล ชื่อ ‘model1’ โดยวิธี centroid พร้อมแสดงความแม่นยำบนชุดข้อมูลทดสอบ .....
15. บันทึกโมเดลในข้อ 14 ในรูปของไฟล์ pickle ตั้งชื่อ face_centroid.pkl
16. สร้างโมเดล ชื่อ ‘model2’ โดยวิธี 1-NN พร้อมเปรียบเทียบผลลัพธ์บนชุดข้อมูลทดสอบกับโมเดล centroid โดย โมเดลที่มีความแม่นยำในการแยกแยะใบหน้าคือ .....