**Lab 5** **204453: Pattern Recognition**

**คำสั่ง:** ให้นักศึกษาดาวน์โหลดไฟล์ชื่อ Iris จาก UC Irvine Machine Learning Repository จากนั้นให้ทำการทดลอง

ขั้นตอนวิธีเพื่อนบ้านใกล้ที่สุดเคตัว (IBk) โดยให้ k มีค่าเท่ากับ 1, 3, 5, 7, 9 โดยรายงานมีรายละเอียด ดังนี้

1. ไฟล์ (ศึกษาจากเอกสารหรือในเว็บไซต์)

* รายละเอียดการรู้จำแบบ (ไฟล์ด้านไหน ต้องการทำนายอะไร)
* ชั้นข้อมูล (แต่ละชั้นคือประเภทอะไร มีความแตกต่างกันอย่างไร)
* ค่าทางสถิติ (จำนวนระเบียน จำนวนสดมภ์ จำนวนชั้นข้อมูล จำนวนข้อมูลในแต่ละชั้นข้อมูล)

1. การกระจายตัวของข้อมูล

* เข้าเมนู Visualize แล้วให้เลือกคู่ของแกนที่เห็นการกระจายตัวชัดเจนที่สุด

1. ผลการทดลองแบบ Cross-validation (Folds = 10) ให้ทดลองแยกแต่ละค่า k

* Accuracy รวมทุกคลาสคำนวณจาก Correctly Classified Instances
* Accuracy แต่ละคลาสคำนวณจาก Confusion Matrix

|  |  |  |  |
| --- | --- | --- | --- |
| **k** | **Accuracy ของ**  **Setosa** | **Accuracy ของ Versicolor** | **Accuracy ของ Virginica** |
| 1 |  |  |  |
| 3 |  |  |  |
| 5 |  |  |  |
| 7 |  |  |  |
| 9 |  |  |  |

1. สรุปผลการทดลอง

* ชั้นข้อมูลใดถูกทำนายแม่นยำมากที่สุด ในแต่ละค่า k
* ชั้นข้อมูลใดถูกทำนายผิดพลาดมากที่สุด ในแต่ละค่า k

1. ทดลองปรับค่าพารามิเตอร์ distanceWeighting และ distanceFunction (อยู่ใน LinearNNSearch)อภิปรายผลการทดลอง ให้นักศึกษาตอบคำถามด้านล่างนี้พร้อมเหตุผลประกอบ

**“พารามิเตอร์ชุดใดให้ค่า Accuracy สูงสุด”**

**หมายเหตุ:** ออกแบบรายงานตามความเหมาะสม ควรให้เรียบร้อย สวยงาม และ อ่านง่าย