

## สรุปหลักการทำงานของ Algorithm ที่ใช้ในการแบ่งบรรทัด (ครั้งที่ 1)

ทำงานโดยการวนลูปของตัวแปร Y ในทุก Segment โดยการลงปากกาไป 1 ครั้งจะนับเป็น 1 Segment และ Y คือจุดของ Y ในแนวแกน Y

โดยในการวนลูปแต่ละครั้งจะมีการตั้งเงื่อนไขในการวนลูปดังนี้

1. เริ่มวนลูปเพื่อหาจุด threshold (จุดที่ไว้ใช้บอกว่ายังอยู่ในบรรทัดนั้นๆหรือไม่)
  - a. วนทั้งหมด 6 รอบในแต่ละรอบจะหาค่า mean (ค่าเฉลี่ย) ของ Segment หรือการลงปากกาในแต่ละครั้ง และเก็บไว้ในตัวแปร Ycmed ซึ่งเป็น list ไว้เก็บค่า mean
  - b. เมื่อวนจนครบ 6 รอบจะได้ list ที่เก็บค่า mean ของ 6 segment มา
  - c. หาส่วนเบี่ยงเบนมาตรฐานของ list ที่เก็บค่า mean เหล่านั้นไว้ในตัวแปร Ystd
  - d. หา mean ของ mean ทั้ง 6 ตัวที่เก็บไว้ในตัวแปร Ycmed และแทนค่า Ycmed ที่เป็นตัวแปรเก็บค่า mean ด้วยค่า mean ทั้งหมดของกลุ่ม เพื่อไว้เป็น threshold สำหรับดูว่ายังอยู่ในบรรทัดนั้นๆหรือไม่
  - e. หาส่วนเบี่ยงเบนมาตรฐานของ list ที่เก็บ mean นั้นไว้เพื่อนำไปคำนวณต่อไป
2. วนลูปเพื่อดูว่าจุดนั้นๆยังอยู่ในบรรทัดนั้นหรือไม่
  - a. หาค่า mean ของ segment ต่อมา
  - b. นำค่า mean ที่หามาได้ไปเปรียบเทียบกับค่า Ycmed โดยมีเงื่อนไขดังนี้
    - i. ถ้า mean ที่ได้  $\geq (Ycmed - 3*Ystd)$  และ  $\leq (Ycmed + 3*Ystd)$  แสดงว่า segment นั้นยังคงอยู่ในบรรทัดนั้นไม่เปลี่ยนไปไหน
    - ii. แต่ถ้า mean ที่ได้  $< (Ycmed - 3*Ystd)$  หรือ  $> (Ycmed + 3*Ystd)$  แสดงว่า segment นั้นไม่ได้อยู่ในบรรทัดนั้นๆอีกต่อไปแล้ว จะเริ่มการหา threshold ใหม่โดยการวนลูปใหม่ให้ segment ที่ไม่ได้อยู่ในบรรทัดเดิมแล้วเป็นตัวเริ่มต้น

## สรุปผล

ผลที่ได้จะออกมาค่อนข้างดีหากมีการเขียนอย่างเป็นระเบียบ และมีการเขียนตัวอักษรต่างๆโดยที่ไม่ห่างกันเกินไป

แต่ผลจะออกมาแย่มากๆ หากว่าตัวอักษรบางตัวที่เขียน มีขนาดค่อนข้างจะต่างกัน หรือมีจุดที่เกิดจากปากกาพลาดไปโดนห่างกันมาก จะเกิดเป็น outlier สำหรับ Segment นั้นๆ และหากเป็นภาษาอังกฤษ จะเป็นกรณีที่ตัวอักษรนั้นๆ มีการขีดในแต่ละครั้งที่อยู่ห่างกันอย่างชัดเจน (eg. T, i) จะทำให้การนำ Y มาหาค่า mean นั้นๆ เกิดเป็น Outlier (ค่าที่ผิดปกติ) และจะเกิดความผิดพลาดเมื่อนำ segment ถัดไปมาเทียบ