1. ข้อใดเป็นคำจำกัดความของการเรียนรู้ด้วยเครื่อง (Machine Learning)

ก. การสอนให้คนเรียนรู้จากคอมพิวเตอร์

ข. การคำนวณเลขด้วยสูตรสมการโดยใช้คอมพิวเตอร์หลายๆเครื่องพร้อมๆกัน

ค. ศาสตร์ที่สอนให้คอมพิวเตอร์เรียนรู้จากข้อมูลโดยที่มนษย์ไม่ต้องโปรแกรม

ง. การประมวลผลข้อมูลภาพ เสียง ภาษา

2. โมเดลในศาสตร์ของการเรียนรู้ด้วยเครื่อง (machine learning model) คือ อะไร

ก. แนวความคิดคณิตศาสตร์พื้นฐานที่เป็นหลักสำคัญของการเรียนรู้ด้วยเครื่อง

ข. สมการทางคณิตศาสตร์ที่นำมาตัดสินว่าคอมพิวเตอร์เราสามารถทำนายข้อมูลได้ถูกต้องแค่ไหน

ค. ตัวแปรทางคณิตศาสตร์ที่เราสร้างขึ้นเพื่อแทนค่าข้อมูลที่เราใช้เป็น input และ output

ง. สมการทางคณิตศาสตร์ที่รับข้อมูลเข้าไปเป็น input และให้ค่าการทำนายออกมาเป็น output

3. โมเดลที่มีลักษณะ overfitting คือโมเดลที่

ก. ถูกคนนำไปใช้เยอะเกินไป จนทำให้ประสิทธิภาพลดลง

ข. จดจำข้อมูลที่ใช้ในการเทรนนิ่งจนทำนายได้ดีแต่ไม่สามารถที่จะทำนายข้อมูลชุดใหม่ที่ไม่เคยเห็นได้

ค. สามารถทำนายข้อมูลได้มีความแม่นยำสูง

ง. มีความซับซ้อนมากจนทำให้คนเข้าใจได้ยาก

4. การใช้ข้อมูลจำนวนมากมาเทรน machine learning model เป็นสิ่งที่ดีหรือไม่ เพราะอะไร

ก. ดี เพราะทำให้โมเดลมีโอกาส overfit น้อยลง

ข. ดี เพราะทำให้คอมพิวเตอร์ช่วยกันคิดได้หลายๆเครื่อง

ค. ไม่ดี เพราะจะทำให้โมเดลมีความแม่นยำน้อยลง

ง. ไม่ดี เพราะอาจจะทำให้โมเดลสับสนได้

5. ข้อใดต่อไปนี้กล่าวผิดเกี่ยวกับ Linear และ Logistic Regression

ก. Linear Regression ใช้สำหรับทำนายข้อมูลที่เป็นตัวเลข (continuous หรือ numeric value)

ข. Logistic Regression ใช้สำหรับทำนายข้อมูลที่เป็นประเภท (categorical หรือ nominal value)

ค. Linear Regression เราใช้ mean square error เป็น cost function

ง. Logistic Regression เราใช้ probabilistic error เป็น cost function

6. Random Forest ต่างจาก Decision Tree อย่างไร

ก. Random Forest คือการเอา Decision Tree หลายๆโมเดลมารวมกันเพื่อช่วยกันตัดสินใจ

ข. Random Forest จะมีการใช้ Random Variable มาช่วยในการทำนายข้อมูล ในขณะที่ Decision Tree จะมีแต่ Static Variable

ค. Decision Tree จะตัดสินใจโดยใช้การแตกกิ่งเพื่อสร้างกฏของการตัดสินใจ ส่วน Random Forest ก็จะมีการแตกกิ่งเช่นกัน แต่กฏมีความซับซ้อนมากกว่า

ง. Random Forest การทำเทรนนิ่งจะยากกว่า Decision Tree เนื่องจากมีโอกาส Overfit ได้สูงกว่า

7. Nearest Neighbor Methods มีสมมติฐานอย่างไรเกี่ยวกับข้อมูล

ก. ข้อมูลมีการกระจุกตัวเป็นกลุ่ม แต่ละกลุ่มแยกตัวจากกลุ่มอื่น

ข. ข้อมูลจุดที่อยู่ใกล้กัน มีคุณลักษณะคล้ายกันจะอยู่คลาสเดียวกัน

ค. ข้อมูลมีความแปรผันกันเป็นเส้นตรง

ง. ข้อมูลมีความหลากหลายและมีการกระจายตัวสูง

8. Supervised Learning ต่างจาก Unsupervised Learning อย่างไร

ก. Supervised Learning ต้องการข้อมูลที่มีการ label แล้วโดยคนหรือมี label จากสิ่งที่เกิดขึ้นในอดีต ส่วน Unsupervised Learning จะไม่ต้องใช้ label

ข. Supervised Learning ต้องมีคนคอยช่วยตรวจความถูกต้องในระหว่างเทรนนิ่ง มิฉะนั้นจะเกิดความผิดพลาดได้ง่ายในขณะที่ Unsupervised Learning โมเดลจะเทรนได้เอง

ค. Unsupervised Learning จะต้องใช้ข้อมูลจำนวนมากกว่า Supervised Learning ในทุกกรณีเพราะโมเดลมักมีความซับซ้อนกว่า

ง. ถูกทุกข้อ

9. เราใช้เทคนิคใดในการเทรนโมเดล Neural Network

ก. Gradient Descent

ข. Hill Climbing Algorithm

ค. Stochastic Optimization

ง. Bayesian Inference

10. Deep Learning มีประโยชน์อย่างไร

ก. สามารถเรียนรู้และเข้าใจข้อมูลภาพได้

ข. สามารถเรียนรู้และเข้าใจข้อมูลเสียงได้

ค. สามารถเรียนรู้และเข้าใจข้อมูลภาษาได้

ง. ถูกทุกข้อ