

Machine Learning

การเรียนรู้ของเครื่อง

Krittameth Teachasrisaksakul

1.1 บทนำ เรื่อง Machine Learning (Introduction to Machine Learning)

Krittameth Teachasrisaksakul

คอร์สนี้จะสอนเกี่ยวกับ

อัลกอริธึมการเรียนรู้ของเครื่อง (Machine Learning Algorithm)

ตัวอย่างการใช้ Machine Learning



- Search engine (เว็บที่ช่วยในการค้นหา) ออนไลน์ เช่น Google กับ Bing ที่ใช้ algorithm จัดอันดับเว็บเพจ
- App ที่ แยกประเภทจากรูปถ่ายเช่น Facebook, Apple ที่ทำการจดจำใบหน้า (face recognition) จากรูปถ่าย
- ตัวกรอง spam แยก spam email ออกจาก email ปกติ
- ทำให้หุ่นยนต์ทำความสะอาดบ้าน โดยเรียนรู้จากตัวอย่างหรือ การสาธิตของคน (learning from demonstrations)

Linear Regression (LR) แบบตัวแปรเดียว

LR ทำนายผลลัพธ์ที่เป็นจำนวนจริง โดยใช้ค่า input (ข้อมูลที่ถูกรับป้อนเข้า)

- การประยุกต์ใช้ LR กับการทำนายราคาบ้าน
- แนวคิดเกี่ยวกับ cost function
- แนะนำวิธี gradient descent สำหรับการเรียนรู้

บทเรียนเพิ่มเติม ทบทวน linear algebra

- ความเข้าใจพื้นฐานเกี่ยวกับ จำเป็นต่อการเรียนคอร์สนี้ทั้งคอร์ส โดยเฉพาะเมื่อเรียนเรื่อง model ที่มีหลายตัวแปร
- ถ้าเข้าใจ linear algebra อยู่แล้ว สามารถข้ามบทเรียนเพิ่มเติมพวกนี้ได้

ทำไม Machine Learning (ML) จึงแพร่หลายมาก ทุกวันนี้?

ML เป็นสาขาวิชาที่เติบโตมาจาก ปัญญาประดิษฐ์ (Artificial Intelligence)

- ถูกพัฒนาเป็น ความสามารถ (capability) ใหม่สำหรับคอมพิวเตอร์

เราเขียนโปรแกรมให้เครื่อง (machine) ทำสิ่งเหล่านี้ ได้หรือไม่?

- **ได้:** งานง่ายๆ เช่น หาเส้นทางที่สั้นที่สุดจากจุด A ไปถึงจุด B
- **ไม่ได้:** งานที่ซับซ้อนและน่าสนใจ เช่น การค้นหาข้อมูลจากเว็บ (web search), การติดแท็กให้รูปภาพ (photo tagging), email anti-spam
- วิธีเดียวที่ทำได้ คือ ให้เครื่อง (machine) เรียนรู้ที่จะทำสิ่งเหล่านี้ด้วยตัวมันเอง

การเติบโตของเว็บและ automation

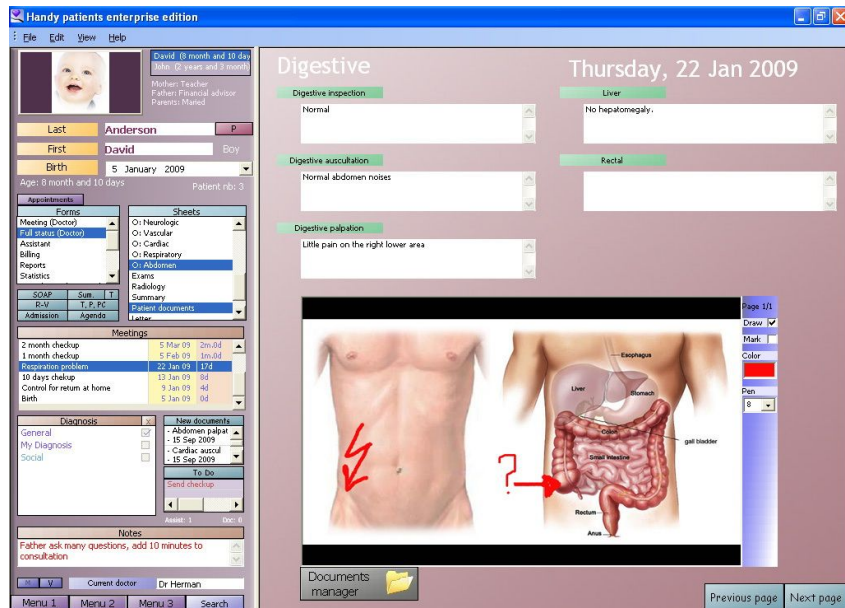
- ซึ่งหมายความว่า เราจะมีชุดข้อมูลขนาดใหญ่มากขึ้น กว่าที่เคยมีมา

ตัวอย่าง: แนวปัญหาที่ใช้ ML แก้



- หุ่นยนต์อัตโนมัติ (autonomous robotics)
- ชีววิทยาเชิงคำนวณ (computational biology)
- ระบบคอมพิวเตอร์ (computer systems)

ตัวอย่าง ML



การทำเหมืองข้อมูล (Database mining)

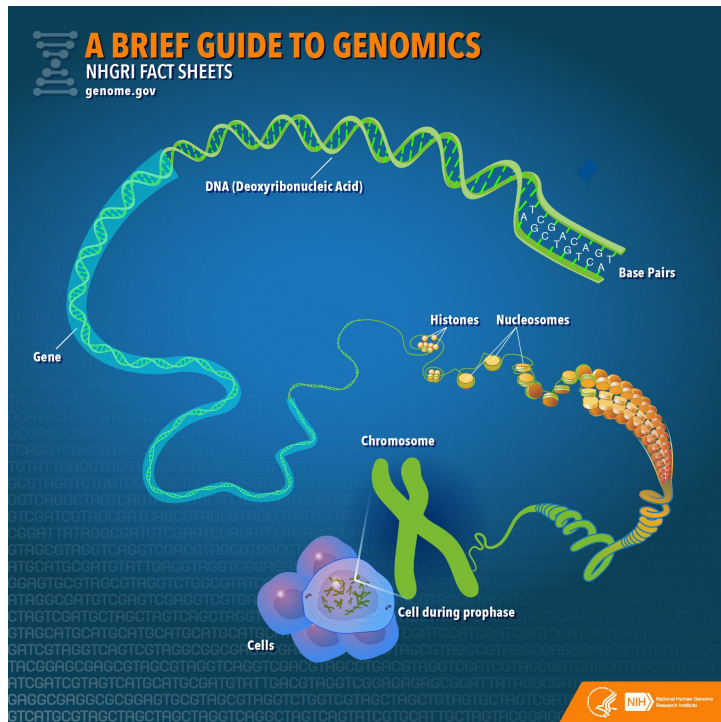
ข้อมูลการคลิกของเว็บ (clickstream data)

- บริษัทที่ Silicon Valley เก็บข้อมูลพวกนี้
- ใช้อัลกอริธึม ML สกัดหรือแยกข้อมูลพวกนี้ เพื่อจะเข้าใจและ
บริการผู้ใช้ได้ดีขึ้น

บันทึกเวชระเบียน แบบอิเล็กทรอนิกส์ (Electronic medical records)

- ถ้าเราสามารถเปลี่ยน บันทึกเวชระเบียน เป็นความรู้ทางการแพทย์ เราจะสามารถเข้าใจโรคได้ดีขึ้น

ตัวอย่าง ML (2)



Computational Biology

- เก็บข้อมูล DNA sequences (ลำดับ DNA = ชุดของอักษรที่แทนนิวคลีโอไทด์)
- run อัลกอริธึมเพื่อให้เข้าใจ genome ของคนมากขึ้น
- genome = ข้อมูลทางพันธุกรรม (เช่น ยีน) ทั้งหมดของสิ่งมีชีวิตหนึ่ง ๆ

วิศวกรรมศาสตร์ (ทุกสาขา)

- มีชุดข้อมูลที่ขนาดใหญ่ขึ้นเรื่อยๆ
- พยายามเข้าใจโดยใช้อัลกอริธึมการเรียนรู้

การประยุกต์ใช้ที่ไม่สามารถป้อนคำสั่ง (program) ได้ด้วยมือ

my alarm clock did not
my alarm code circle shute rout
clock risk riot not
did visit must

wake me up this morning
wake me up thai taxi moving
this tier having
morning loving

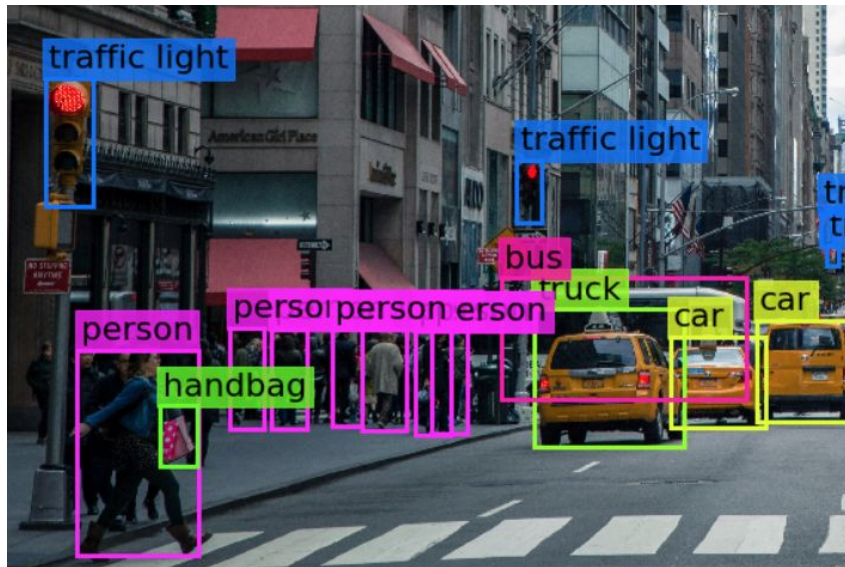
เฮลิคอปเตอร์ที่ทำงานด้วยตนเอง (Autonomous Helicopter)

- ให้คอมพิวเตอร์เรียนรู้ด้วยตนเองว่าจะควบคุมเฮลิคอปเตอร์ให้บินได้อย่างไร

การจดจำลายมือเขียน (Handwriting Recognition)

- อัลกอริธึมการเรียนรู้สามารถอ่านลายมือเขียนเพื่อจัดส่งจดหมายให้ถึงที่อยู่ปลายทาง
- เพื่อลดต้นทุนการส่งจดหมาย

การประยุกต์ใช้ที่ไม่สามารถป้อนคำสั่ง (program) ได้ด้วยมือ (2)



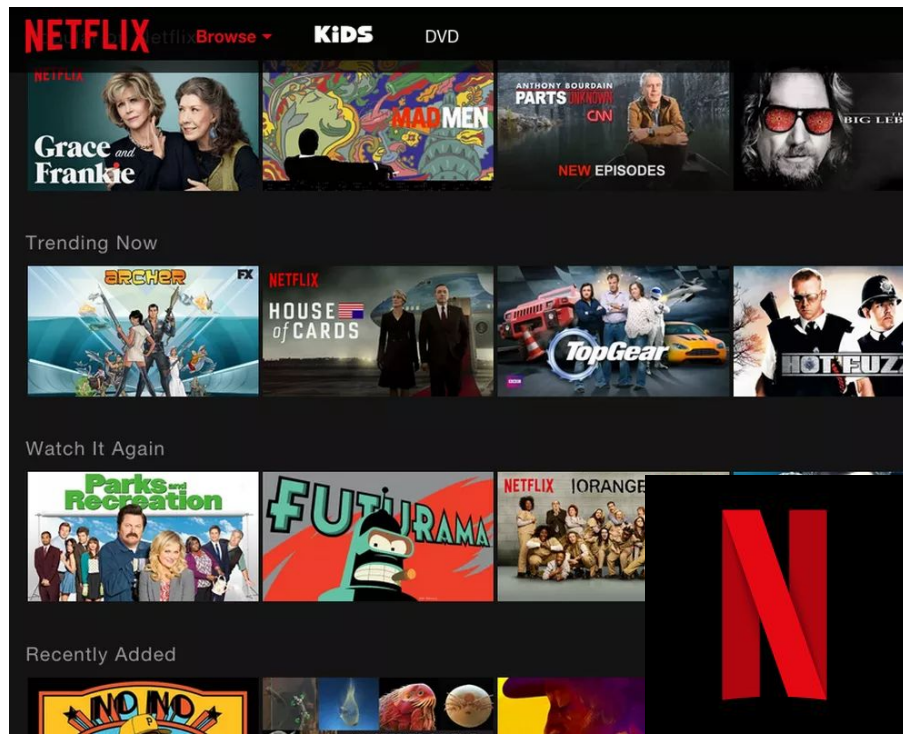
การประมวลผลภาษาธรรมชาติ (Natural Language Processing : NLP)

= สาขาวิชาของ AI เกี่ยวกับการเข้าใจภาษา

คอมพิวเตอร์วิทัศน์ (Computer Vision : CV) = สาขาวิชาของ AI เกี่ยวกับการเข้าใจรูปภาพ

- การประยุกต์ด้าน NLP และ CV ส่วนมากเป็น machine learning แบบประยุกต์

การประยุกต์ใช้ที่ไม่สามารถป้อนคำสั่ง (program) ได้ด้วยมือ (3)



โปรแกรมที่ **customize** (ปรับแต่งตามความต้องการของผู้ใช้) ได้ด้วยตนเอง เช่น Amazon, Netflix, iTunes Genius playlist

- ใช้อัลกอริธึมการเรียนรู้ แนะนำผลิตภัณฑ์ (product recommendation) เช่น ภาพยนตร์ กับ เพลง ให้กับผู้ใช้
- เพราะมีลูกค้าเป็นหลักล้านคน วิธีเดียวที่จะทำให้ software ให้คำแนะนำแบบเฉพาะบุคคลได้ คือ ให้มันเรียนรู้ด้วยตนเองที่จะ customize ตัวมันให้ตรงกับความชื่นชอบของผู้ใช้แต่ละคน

1.2 การเรียนรู้ของเครื่อง (Machine Learning: ML) คืออะไร?

Krittameth Teachasrisaksakul

หัวข้อที่จะเรียน

- นิยามของ Machine Learning (ML)
- ตัวอย่างสถานการณ์ที่ใช้ ML ได้

Machine Learning (การเรียนรู้ของเครื่อง) : นิยาม (1)

Arthur Samuel (1959)

- ได้ให้นิยามที่เก่าและไม่เป็นทางการ (informal)

"Machine Learning: Field of study that gives computer the ability to learn without being explicitly programmed."

(แปล) "Machine Learning คือ สาขาวิชาที่ทำให้ คอมพิวเตอร์ สามารถเรียนรู้ (จากตัวอย่าง/ข้อมูล) ได้ด้วยตนเอง โดยโปรแกรมเมอร์ไม่ต้องป้อนคำสั่ง" [2]

Arthur Samuel



- โด่งดังเรื่องเขียนโปรแกรม (ป้อนคำสั่ง) ให้คอมพิวเตอร์เล่น checkers (หมากฮอส)
- ป้อนคำสั่งให้คอมพิวเตอร์เล่น checkers แข่งกับตัวเอง นับหมื่นครั้ง
- โปรแกรมเรียนรู้ตำแหน่งที่วางหมากที่ดีและไม่ดี จากการดูตำแหน่งการวางหมากที่ทำให้ชนะและแพ้
- ท้ายที่สุด โปรแกรมเรียนรู้ที่จะเล่น checkers ได้ดีกว่า Arthur Samuel

Machine Learning (การเรียนรู้ของเครื่อง) : นิยาม (2)

Tom Mitchell (1998) : ได้ให้คำนิยามที่ทันสมัยกว่า

"Well-posed Learning Problem: A computer program is said to *learn* from experience E with respect to some task T and some performance measure P , if its performance on T , as measured by P , improves with experience E ."

- (แปล) "เราบอกได้ว่า computer program “เรียนรู้” จากประสบการณ์ (หรือข้อมูล/ตัวอย่าง) E เมื่อประสิทธิภาพของ T ที่ถูกวัดด้วย P ดีขึ้น (ทำงานได้ดีขึ้น)"
 - T = task = งาน/ภารกิจ; P = performance measure = ตัววัดประสิทธิภาพ
- "computer program ที่ถูกพัฒนาขึ้นเพื่อเอาไปทำงานใดงานหนึ่ง เมื่อมีข้อมูลป้อนให้ ถ้า program นั้นทำงานได้ดีขึ้น ก็หมายความว่ามันได้ “เรียนรู้” ที่จะทำงานนั้นๆ แล้ว" [3]

Machine Learning (การเรียนรู้ของเครื่อง) : นิยาม (2)

ตัวอย่าง: การเล่นเกมหมากฮอส (checkers)

- E = ประสบการณ์ (experience) หรือตัวอย่างของการเล่นหลายๆครั้ง แข่งกับตัวเอง
- T = สิ่งที่ทำ, งาน, หรือภารกิจ (task) ในการเล่นเกมหมากฮอส
- P = ความน่าจะเป็น (probability) ที่โปรแกรมจะชนะในการเล่นต่อไป แข่งกับคู่ต่อสู้คนใหม่

Quiz: สถานการณ์ตัวอย่าง คือ โปรแกรมที่บอกว่า email แต่ละอันเป็น spam หรือไม่ โดยเรียนรู้จากข้อมูลว่า email ไหนที่เราบอกว่าเป็น spam

ให้ใช้คำนิยามของ Machine Learning โดย Tom Mitchell เพื่อตอบว่า E (Experience), T (Task), P (Performance Measure) ในสถานการณ์นี้ คืออะไร?

Quiz: สถานการณ์ตัวอย่าง คือ โปรแกรมที่บอกว่า email แต่ละอันเป็น spam หรือไม่ โดยเรียนรู้จากข้อมูลว่า email ไหนที่เราบอกว่าเป็น spam

ให้ใช้คำนิยามของ Machine Learning โดย Tom Mitchell เพื่อตอบว่า E (Experience), T (Task), P (Performance Measure) ในสถานการณ์นี้ คืออะไร?

Answer: ในสถานการณ์นี้

- **Task T** คือ การแยกประเภทของอีเมล : เป็น spam, ไม่เป็น spam
- **Experience E** คือ การดูว่าเราทำเครื่องหมายว่าอีเมลเป็น spam หรือไม่เป็น spam
- **Performance measure P** คือ จำนวนหรือสัดส่วนของ email ที่ถูกแยกประเภท (ว่าเป็น spam หรือไม่เป็น spam) ได้ถูกต้อง
- ประสิทธิภาพของระบบเราในการทำ task T (โดยใช้ performance measure P เป็นตัววัด) จะดีขึ้นหลังจากมีประสบการณ์ E
- ก็คือ ระบบหรือโปรแกรมจะทำ task T ได้ดีขึ้น เมื่อมันเรียนรู้จากตัวอย่าง หรือ experience E

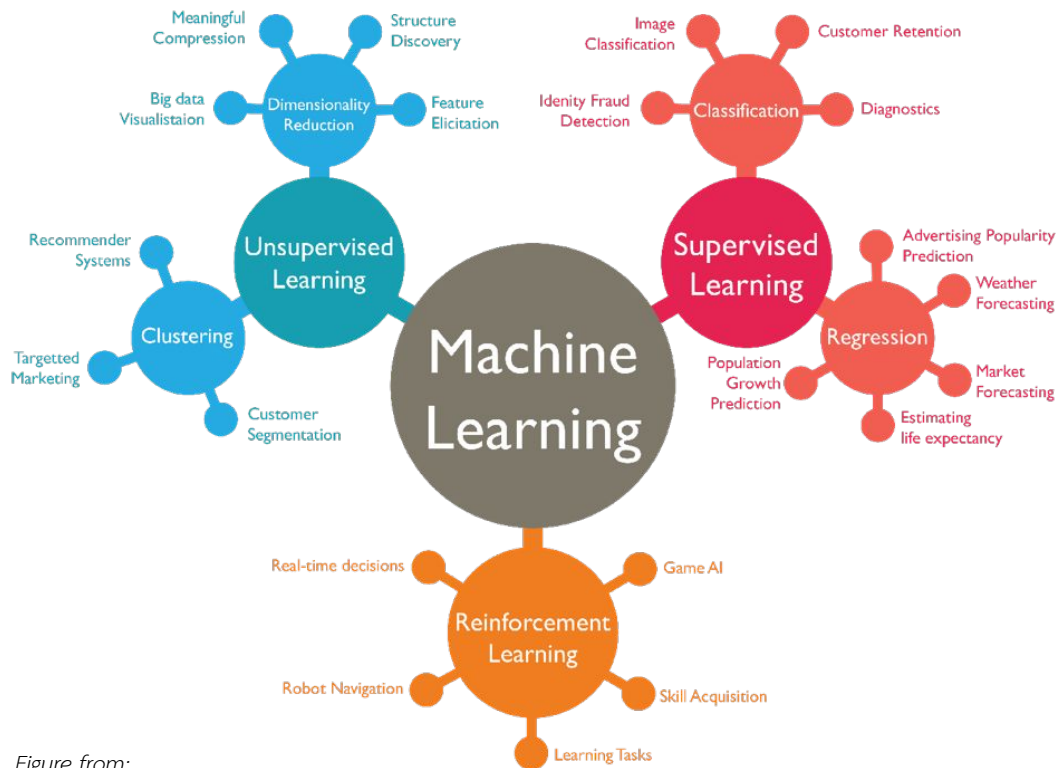
1.3 Supervised Learning (SL)

Krittameth Teachasrisaksakul

หัวข้อที่จะเรียน

- ประเภทของอัลกอริธึมการเรียนรู้ (learning algorithms) 2 ประเภท
- คำแนะนำที่นำไปใช้ได้จริง สำหรับประยุกต์ใช้ learning algorithms
- **แนวปฏิบัติที่ดี (Best practices)** ในด้าน Machine Learning (ML) และ Artificial Intelligence (AI)
- การตัดสินใจเกี่ยวกับวิธีที่ใช้สร้างระบบ ML
 - เพื่อเพิ่มโอกาสประสบความสำเร็จเมื่อเราประยุกต์ใช้ learning algorithm
 - หลีกเลี่ยงการเสียเวลาศึกษาค้นคว้าบางหัวข้อไป 6 เดือน แล้วพบว่ามันทำงานไม่ได้
- วิธีทำให้ ML และ AI ทำงานได้
- เป้าหมาย / จุดประสงค์การเรียนรู้: รู้วิธีสร้างระบบ ML และ AI

ประเภทของอัลกอริทึมการเรียนรู้ (Learning algorithms)

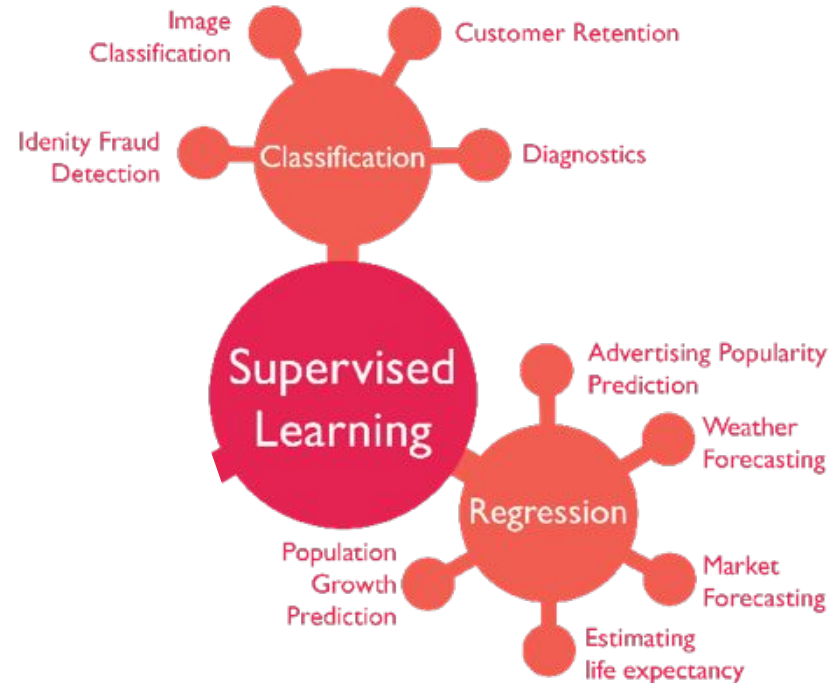


- **Supervised Learning:**
สอนคอมพิวเตอร์
ทำงานบางอย่าง
- **Unsupervised Learning:**
ให้คอมพิวเตอร์
เรียนรู้ด้วยตนเอง
- **Reinforcement Learning, Recommender Systems**

Figure from:

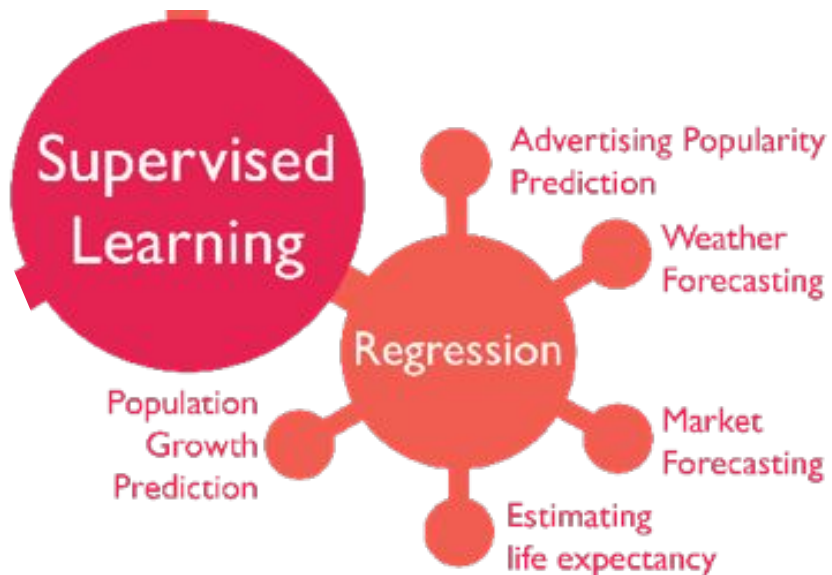
<https://blogs.oracle.com/ai/types-of-machine-learning-and-top-10-algorithms-ever-you-should-know>

1. Supervised Learning (SL)



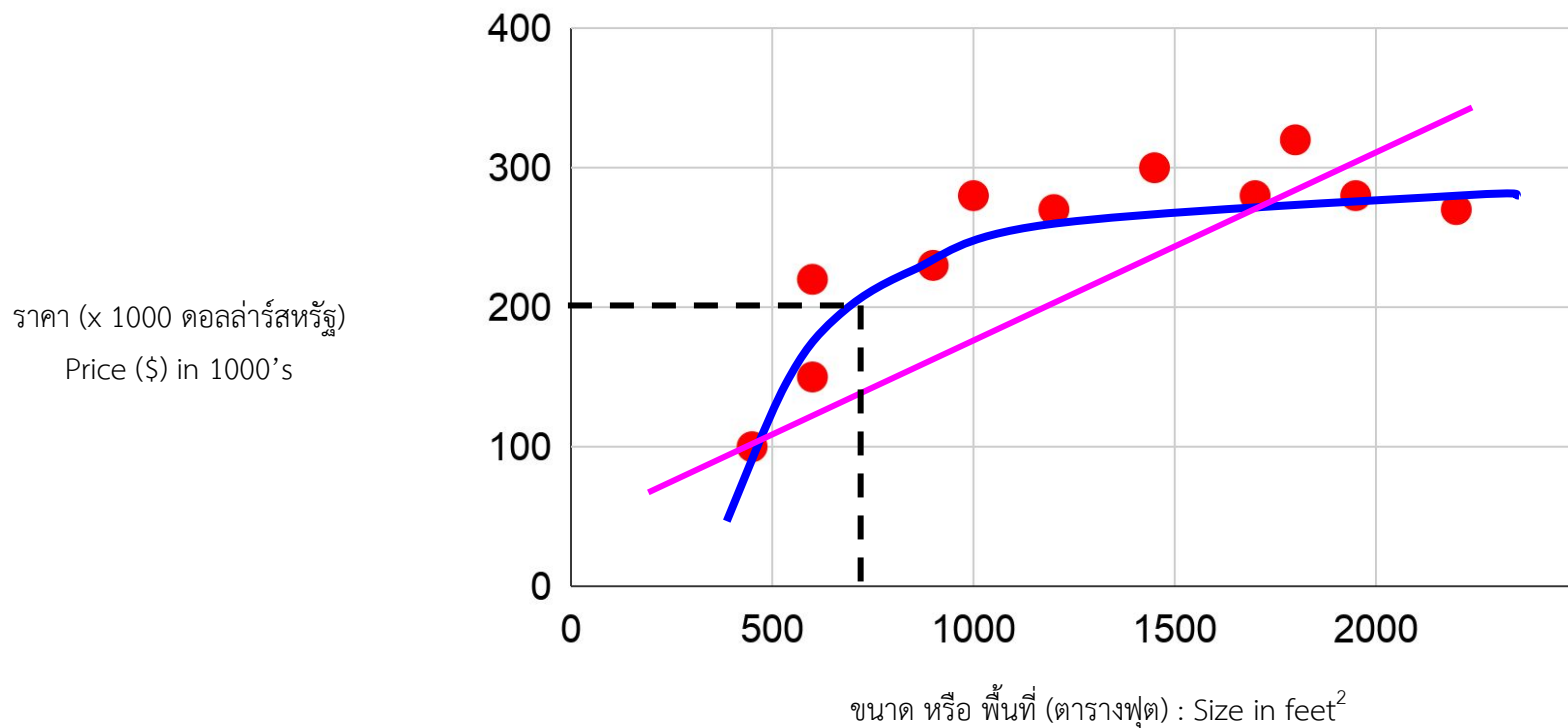
- ประเภทของปัญหา Machine Learning ที่พบได้ทั่วไปมากที่สุด
- ใช้แก้ปัญหาที่เรา **รู้** คำตอบ / ผลลัพธ์ (output) ที่ถูกต้อง และ ส่งชุดข้อมูลให้กับ algorithm
- เช่น ชุดข้อมูลบ้าน และบอกราคาจริงของบ้าน แต่ละหลัง

1.1 Regression Problem



- ทำนายค่าผลลัพธ์ที่ต่อเนื่อง (continuous value) (ที่เป็นจำนวนจริง, scalar)
- เช่น ทำนายราคาบ้าน, ทำนายอายุคน จากรูปถ่ายของคน

ตัวอย่าง SL (1): ทำนายราคาบ้าน



ตัวอย่าง SL (1): ทำนายราคาบ้าน

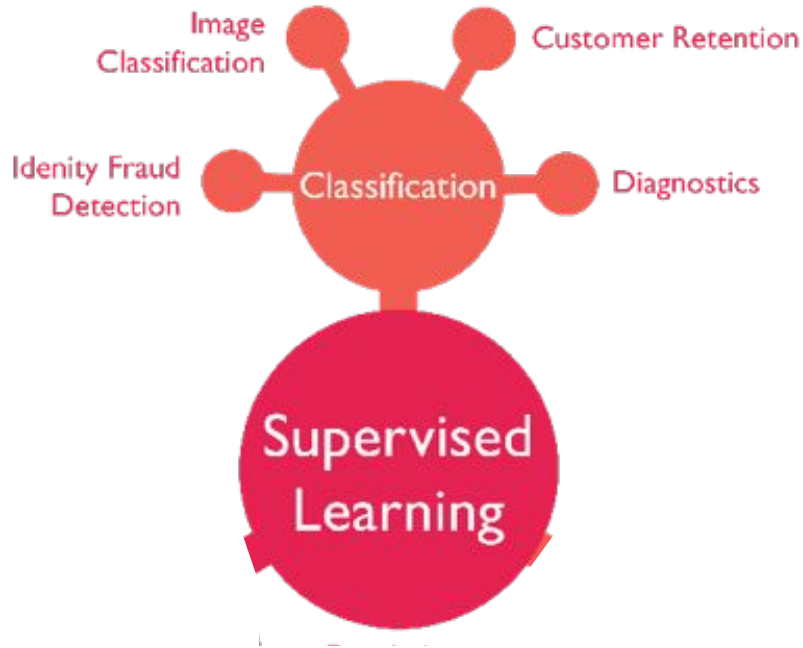
ดู plot ของชุดข้อมูล: ราคาบ้าน (1000 USD) กับ ขนาดบ้าน (ตารางฟุต หรือ feet^2)

ปัญหา: ทำนายราคาของบ้านขนาด 750 ตารางฟุต เช่น เพื่อนอยากขายบ้าน และอยากรู้ราคาตลาด

- วิธี 1: ใช้ learning algorithm เพื่อใช้เส้นตรงทำนายชุดข้อมูล (จุดข้อมูล)
 - ได้ราคาประมาณ 150,000 USD
- วิธี 2: ใช้ฟังก์ชันกำลังสอง / quadratic function (พหุนามกำลังสอง) ทำนายข้อมูล
 - ได้ราคาประมาณ 200,000 USD

เราจะพูดถึงวิธีเลือก / ตัดสินใจว่าจะใช้วิธีไหน (เส้นตรง หรือ quadratic function) เพื่อแก้ปัญหาการทำนายราคาบ้าน

1.2 ปัญหาการแยกประเภท (Classification Problem)



ทำนายค่าผลลัพธ์ที่ไม่ต่อเนื่อง (discrete value) ที่มีมากกว่าหรือเท่ากับ 2 ค่า เช่น

- {0,1}: เนื้องอก {ไม่ร้าย, ร้าย} หรือ {benign, malignant}
- {0,1,2,3}: {ไม่ร้าย (ไม่เป็นมะเร็ง), มะเร็งชนิดที่ 1, มะเร็งชนิดที่ 2, มะเร็งชนิดที่ 3} หรือ {benign (no cancer), type I cancer, type II cancer, type III cancer}

ตัวอย่าง SL (2): ทำนายการเป็นโรคมะเร็งเต้านม (Breast cancer)

- ดูข้อมูลจากเวชระเบียนหรือบันทึกรายงานผู้ป่วย (medical records) เกี่ยวกับเนื้องอก (tumor) ในเต้านม (ก้อนนูนในเต้านม)
- ทำนายว่า เนื้องอกเป็นเนื้องอก ร้ายแรง (malignant / เป็นอันตราย) หรือ ไม่ร้าย (benign / ไม่อันตราย)
- plot ของขนาดเนื้องอก (แกนราบ / แกน x) vs เป็นเนื้องอกร้ายหรือไม่ มีแค่ค่า 1 หรือ 0 (แกนตั้ง / แกน y)
- สมมติ คนไข้มีเนื้องอกในเต้านม และเรารู้ขนาดเนื้องอก

คำถาม ML: ให้เราประมาณค่า ความน่าจะเป็น หรือโอกาสที่เนื้องอกจะร้ายแรง (malignant) หรือ ไม่ร้าย (benign)

ตัวอย่างนี้เป็นปัญหาการแยกประเภท (classification problem)

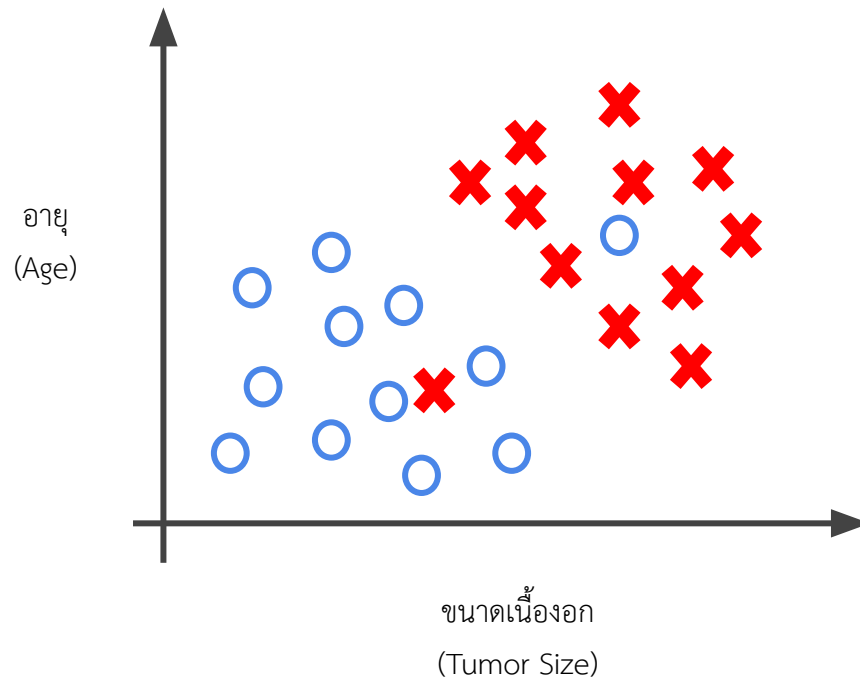
ตัวอย่าง SL (2): ทำนายการเป็นโรคมะเร็งเต้านม (Breast cancer)

สมมติ เรารู้อายุและขนาดเนื้องอก ของคนไข้

- ใน plot ให้
- ○ แทน คนไข้ที่มีเนื้องอกไม่ร้าย (benign tumor)
- × แทน คนไข้ที่มีเนื้องอกร้าย (malignant tumor)

สมมติ มีคนไข้คนหนึ่งมีเนื้องอก

ข้อมูลของคนไข้ (ขนาดเนื้องอก, อายุ) แทนด้วย จุดน้ำเงิน 1 จุด



ตัวอย่าง SL (2): ทำนายการเป็นโรคมะเร็งเต้านม (Breast cancer)

Learning algorithm อาจใช้เส้นตรงเพื่อแบ่งแยก ชนิด (class) ของเนื้องอก 2

ชนิด: ร้ายแรง, ไม่ร้าย

- ในกรณีนี้ algorithm จะบอกว่าเนื้องอกของคนไข้อยู่ในฝั่ง**ไม่ร้าย** และเป็นไปได้มากกว่าที่จะไม่ร้ายมากกว่าร้ายแรง

ตัวอย่างนี้ มีลักษณะเฉพาะ (feature) 2 อัน: อายุคนไข้ (Age) กับ ขนาดเนื้องอก (Tumor Size)

- ในปัญหา Machine Learning อื่นๆ เรามักจะมี feature จำนวนมากกว่านี้
- คนที่ทำงานกับปัญหานี้จริงๆ จะใช้ feature อื่นๆ ด้วย เช่น ความหนาของก้อนเนื้อ (clump thickness), ความสม่ำเสมอ (uniformity) ของขนาดเซลล์, ความสม่ำเสมอของรูปร่างเซลล์

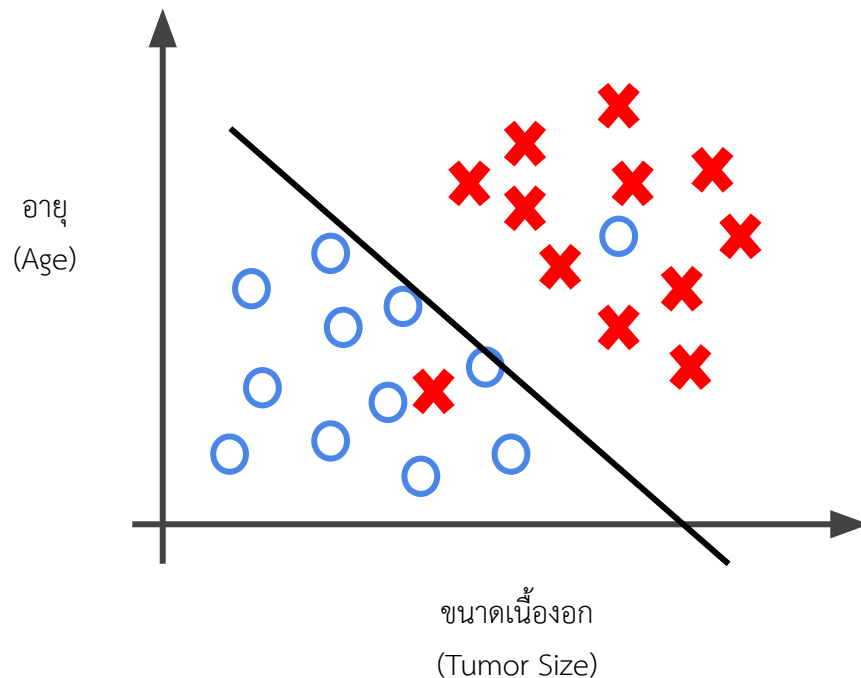


Figure source: Andrew Ng's Machine Learning course

Recap: สรุป

Supervised Learning

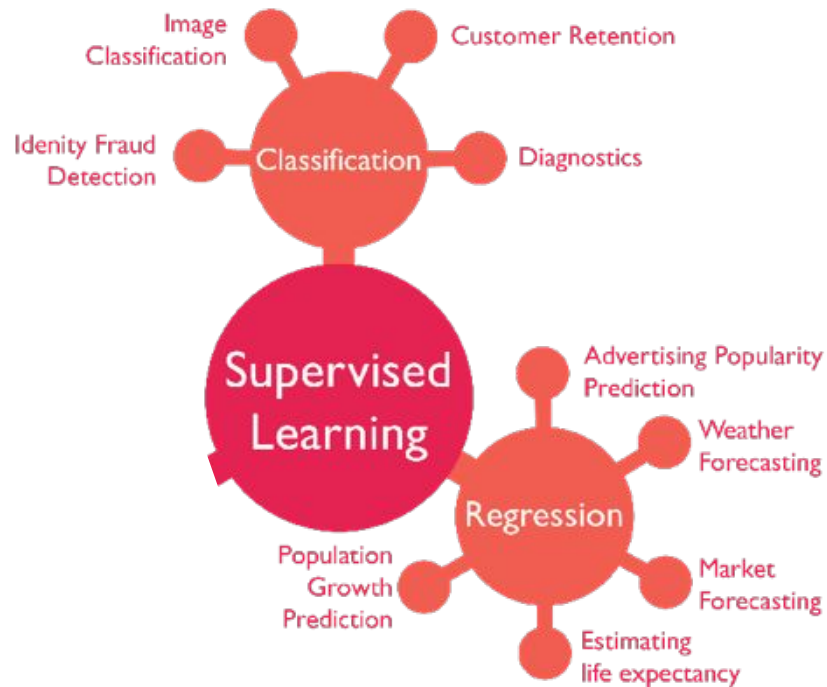
- ทุกๆตัวอย่างในชุดข้อมูล มีคำตอบที่ถูกต้อง
ที่เราอยากให้ algorithm ทำนายเป็นผลลัพธ์ของตัวอย่างนั้นๆ

Regression Problem

- เป้าหมาย: ทำนายผลลัพธ์ที่มีค่าต่อเนื่อง
(continuous value output) เช่น ราคาก่อน

Classification Problem

- เป้าหมาย: ทำนายผลลัพธ์ที่มีค่าไม่ต่อเนื่อง (discrete value output)
- เช่น ประเภทเนื้ออก: ร้าย หรือ ไม่ร้าย



คำถามทบทวน

สมมติ เราทำบริษัทและอยากพัฒนา algorithm ที่แก้ไขปัญหา 2 ปัญหา

- **ปัญหาที่ 1:** เรามี inventory / stock (สินค้าคงคลัง) ที่ใหญ่ มีของที่เหมือนกันเป็น 1000 ชิ้น
เราอยากทำนายว่ามีของกี่ชิ้นที่จะขายได้ใน 3 เดือนถัดไป
- **ปัญหาที่ 2:** อยากให้ software ตรวจสอบบัญชีของลูกค้าแต่ละคน
และบอกว่าบัญชีแต่ละบัญชีถูก hack หรือไม่
- ปัญหาแต่ละอันเป็น classification หรือ regression?

คำถามทบทวน

สมมติ เราทำบริษัทและอยากพัฒนา algorithm ที่แก้ไขปัญหา 2 ปัญหา

- **ปัญหาที่ 1:** เรามี inventory / stock (สินค้าคงคลัง) ที่ใหญ่ มีของที่เหมือนกันเป็น 1000 ชิ้น
เราอยากทำนายว่ามีของกี่ชิ้นที่จะขายได้ใน 3 เดือนถัดไป
- **ปัญหาที่ 2:** อยากให้ software ตรวจสอบบัญชีของลูกค้าแต่ละคน
และบอกว่าบัญชีแต่ละบัญชีถูก hack หรือไม่
- ปัญหาแต่ละอันเป็น classification หรือ regression?

คำตอบ

- ปัญหาที่ 1 เป็นปัญหา regression และ ปัญหาที่ 2 เป็น classification เพราะ ...
- ปัญหาที่ 1 มีจำนวนสิ่งของที่ขายได้ ซึ่งเป็นค่าต่อเนื่อง
- ปัญหาที่ 2 เป็นค่าที่ไม่ต่อเนื่อง ให้ 0 แทน บัญชีไม่ถูก hack และ 1 แทน บัญชีถูก hack

1.4 Unsupervised Learning

Krittameth Teachasrisaksakul

Unsupervised Learning (UL)

เปรียบเทียบ: Supervised Learning (SL) vs. Unsupervised Learning (UL)

- SL: รู้คำตอบ/ผลลัพธ์ (output) /label ที่ถูกต้อง
- UL: มีชุดข้อมูล แต่ไม่มี label

Unsupervised Learning : มีชุดข้อมูล หาโครงสร้างในข้อมูล

- เช่น clustering algorithm แยกข้อมูลเป็นกลุ่ม (cluster)
- clustering เป็น UL ประเภทหนึ่ง

ตัวอย่าง: Clustering

Google News: <https://news.google.com>

- ใช้ clustering
- ดูข่าวใหม่ (ประมาณ 1 แสนข่าว) จากบนเว็บ
- จัดกลุ่มข่าว เป็นเรื่องราวข่าวที่ต่อเนื่อง
- ผลที่ได้ : ข่าวเกี่ยวกับหัวข้อเดียวกันจะถูกแสดงด้วยกัน
- เช่น ข่าวเกี่ยวกับหัวข้อเดียวกัน : BP Oil Well Spill จากแหล่งข่าว Wall Street Journal, CNN, UK Guardian



Headlines

[More Headlines](#)

Whistleblower alleges White House coverup: Live updates

CNN · 11 minutes ago

- White House Knew of Whistle-Blower's Allegations Soon After Trump's Call With Ukraine Leader

The New York Times · 11 hours ago



การประยุกต์ใช้ clustering: (1) ทำความเข้าใจ genomics

- genomics = การศึกษาจีโนม (genome) ทั้งหมดของสิ่งมีชีวิต
- genome = ข้อมูลทางพันธุกรรม (เช่น ยีน) ทั้งหมดของสิ่งมีชีวิตหนึ่ง ๆ
- ตัวอย่างข้อมูล DNA microarray (= ชุดของจุด DNA ขนาดเล็ก ที่ยึดติดกับพื้นผิวที่เป็นของแข็ง)
- จัดกลุ่มของคน โดยวัดว่าแต่ละคนมี หรือไม่มี ยีนบางตัว ก็คือ ยีนถูกแสดงออกมากแค่ไหน
 - สี (แดง, เขียว, เทา, ...) แสดงว่า ยีนถูกแสดงออกมากแค่ไหน
- run clustering algorithm เพื่อจัดกลุ่ม/แบ่งคน เป็นประเภท (ที่เราไม่รู้ล่วงหน้า) โดยอัตโนมัติ
 - Clustering เป็น Unsupervised Learning เราจึงไม่ได้บอก algorithm ล่วงหน้าว่าแต่ละคนเป็นประเภทไหน

การประยุกต์ใช้ clustering: (2) จัดระเบียบ computer cluster ขนาดใหญ่



คนที่บำรุงรักษาและจัดการ data center (หรือ computer cluster) ขนาดใหญ่

- Computer cluster = กลุ่มของคอมพิวเตอร์ที่เชื่อมต่อกันและทำงานด้วยกันเสมือนว่าเป็นเครื่องคอมพิวเตอร์เพียง 1 เครื่อง [4]

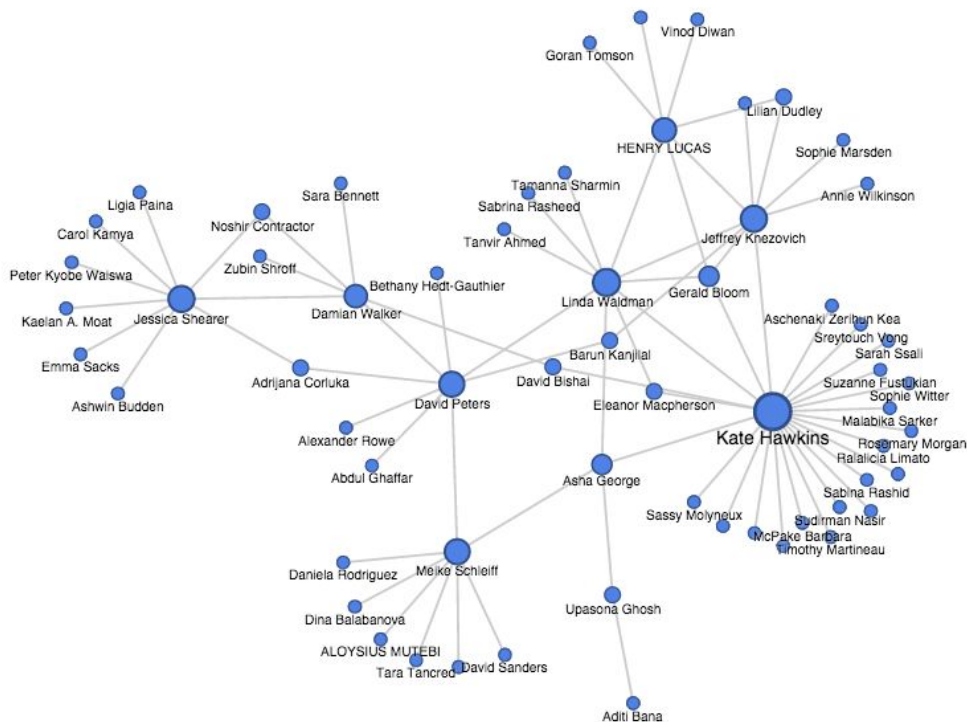
หาว่าเครื่องไหนที่มักจะทำงานร่วมกัน

จัดให้เครื่องเหล่านั้นมาอยู่ด้วยกัน เพื่อเพิ่มประสิทธิภาพการทำงานของ data center

Source:

<https://www.breeze-software.com/services/high-speed-modeling/cluster-computing/>

การประยุกต์ใช้ clustering: (3) การวิเคราะห์เครือข่ายทางสังคม (Social Network)



ให้รายชื่อเพื่อนใน Facebook หรือ ข้อมูลที่เราส่งอีเมลหาเพื่อนคนไหน
บ่อยที่สุด

ให้หากลุ่มคนที่รู้จักกันและกันทั้งหมด โดยอัตโนมัติ

Source:

<http://www.futurehealthsystems.org/blog/2014/10/1/people-centred-health-systems-research-methods>

การประยุกต์ใช้ clustering: (4) การแบ่งส่วนตลาด (Market segmentation)



ให้ฐานข้อมูลขนาดใหญ่เกี่ยวกับข้อมูลลูกค้าของบริษัท

ให้หาส่วนตลาด (market segment) โดยอัตโนมัติ และจัดกลุ่มข้อมูลลูกค้าเป็น market segment ต่างๆ

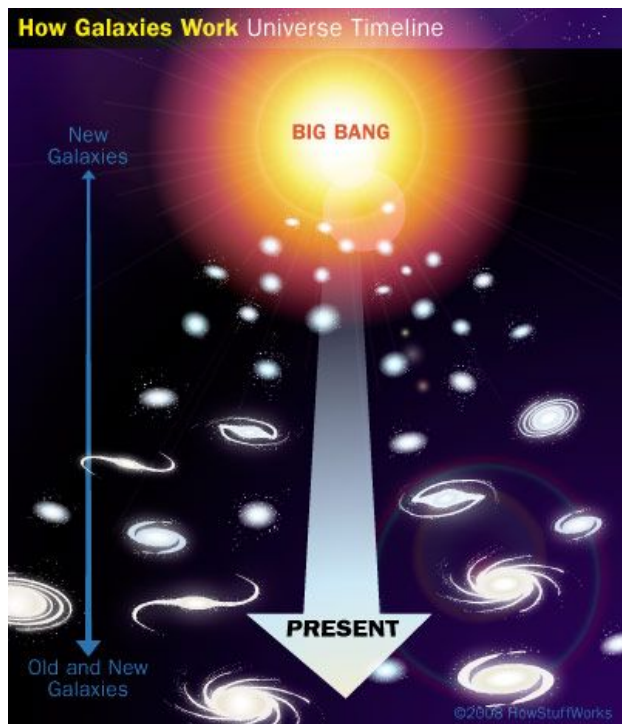
เป้าหมาย: ทำการขายหรือทำการตลาดสำหรับ market segment ต่างๆ โดยอัตโนมัติ และอย่างมีประสิทธิภาพมากขึ้น

ปัญหานี้เป็น UL เพราะเราไม่รู้ว่ามี market segment อะไรบ้าง และลูกค้าแต่ละคนอยู่ใน segment ไດบ้าง

Source:

<https://www.singlegrain.com/digital-marketing/strategists-guide-marketing-segmentation/>

การประยุกต์ใช้ clustering: (5) การวิเคราะห์ข้อมูลทางดาราศาสตร์

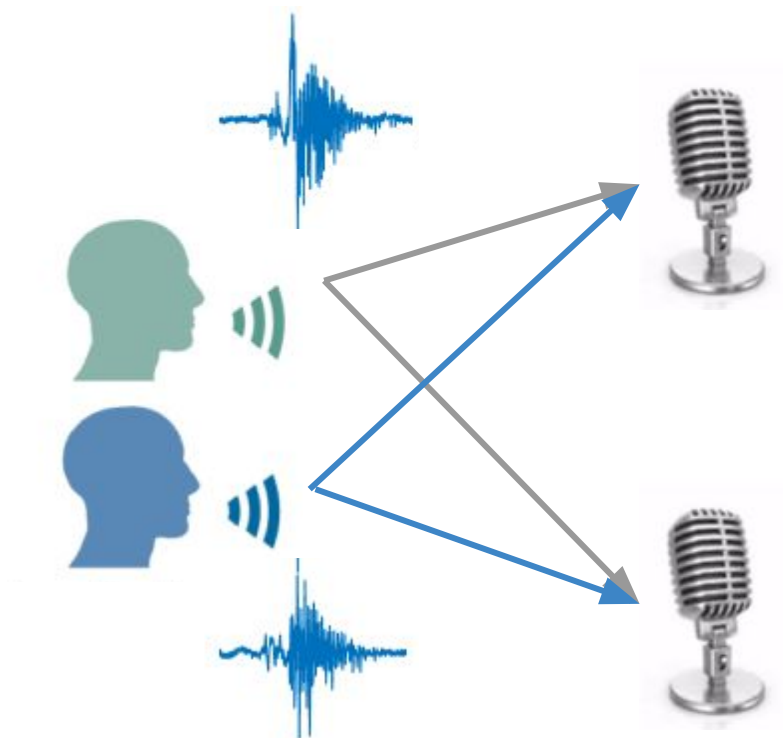


เพื่อหาทฤษฎีที่น่าสนใจและมีประโยชน์เกี่ยวกับ galaxies ก่อตัวขึ้นมา
อย่างไร

Source:

<https://science.howstuffworks.com/dictionary/astronomy-terms/galaxy3.htm>

ปัญหา Cocktail Party



เวลาปาร์ตี้ มีคนเต็มห้อง ทุกคนกำลังพูดพร้อมๆกัน

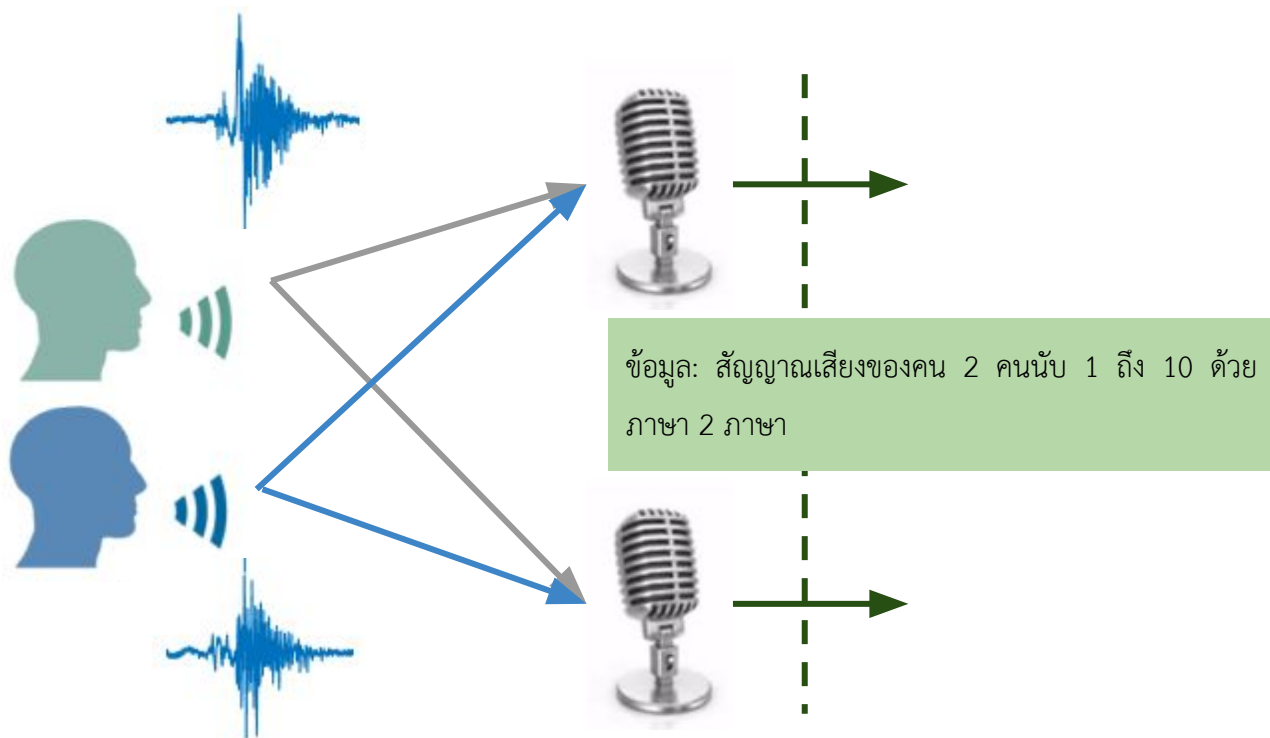
ดังนั้น จะมีเสียงที่ซ้อนทับกัน และได้ยินเสียงคนพูดแต่ละคนไม่ชัด

สมมติว่าเป็น cocktail party ที่มีคน 2 คน และสองคนนี้พูดพร้อมๆกัน

วางไมโครโฟน 2 ตัวในห้อง ที่ระยะ 2 ระยะที่ต่างกันจากคนพูด

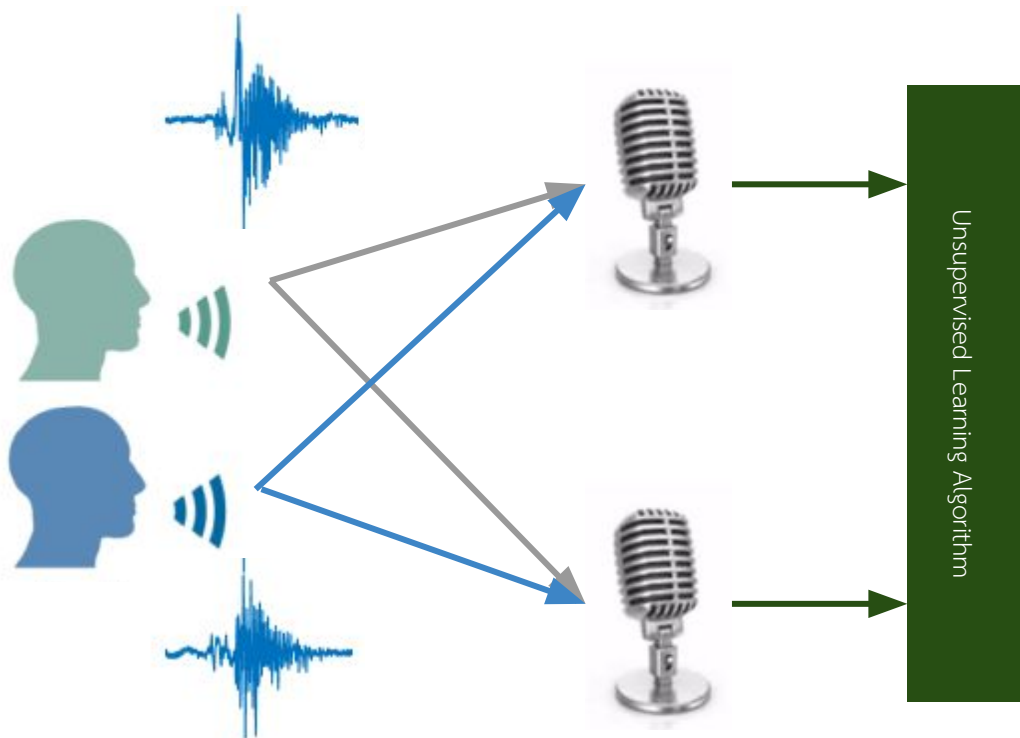
ไมโครโฟนแต่ละตัวอัดเสียงที่ซ้อนทับกันของคนพูด 2 คน

ปัญหา Cocktail Party: ตัวอย่างที่ 1



Source: <https://futurelearning.ai/how-to-solve-the-cocktail-party-problem/>

ปัญหา Cocktail Party: ตัวอย่างที่ 1

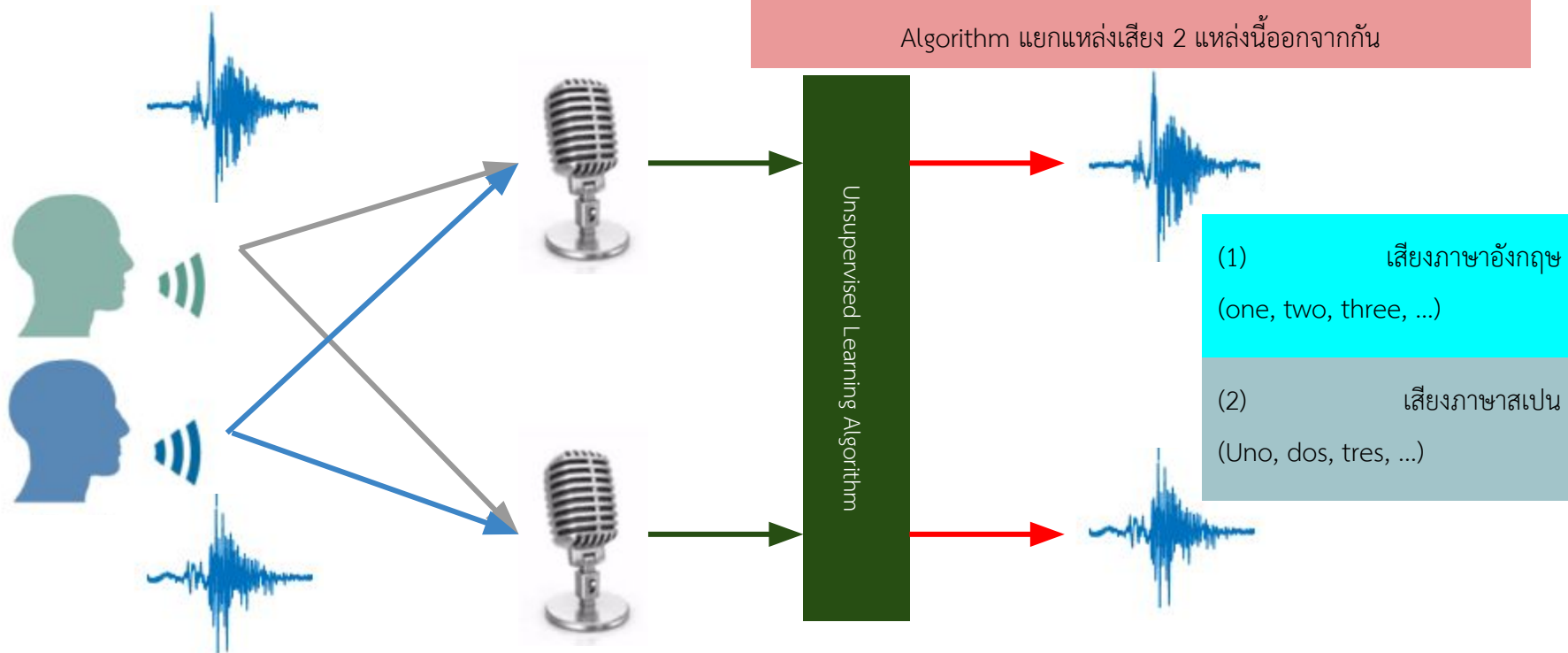


ส่งข้อมูลเหล่านี้ให้กับ cocktail party algorithm
(unsupervised learning algorithm)

ให้ algorithm หาโครงสร้างในข้อมูลนี้

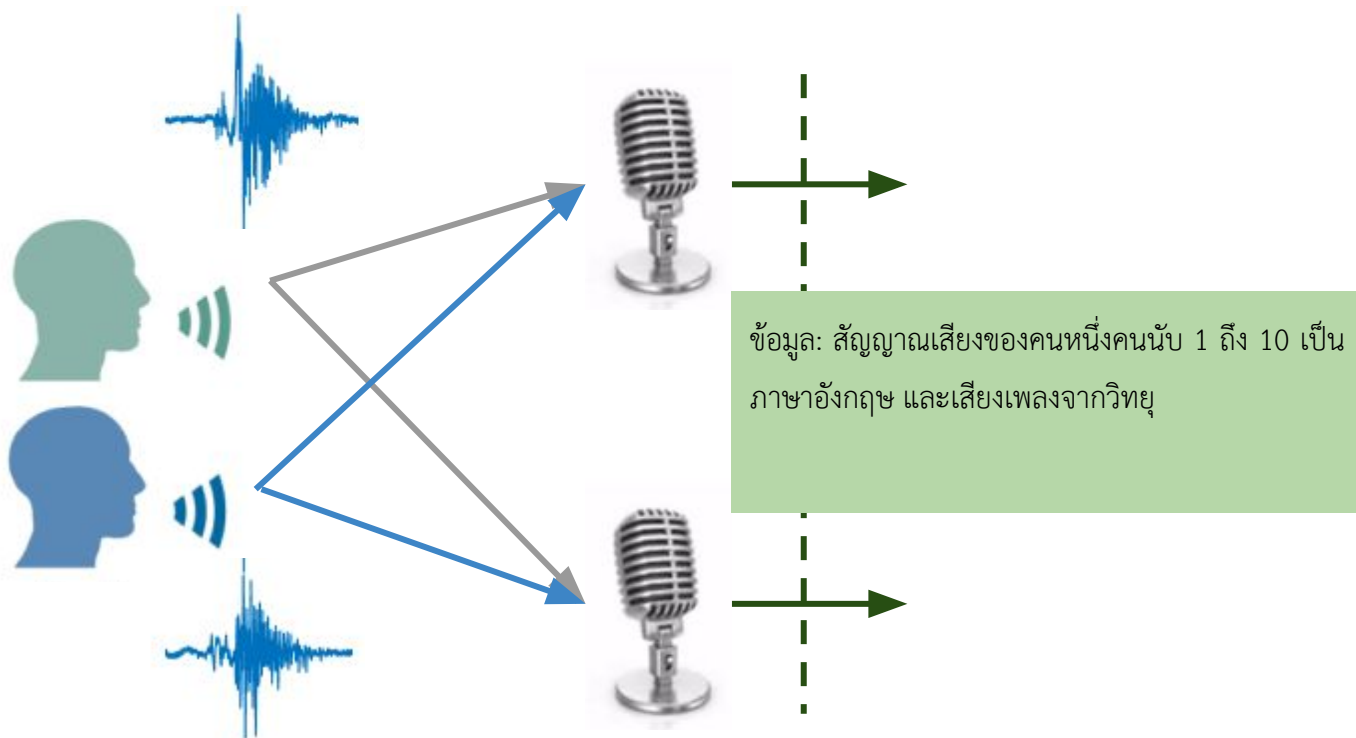
แนวคิดของ algorithm: สมมติว่า ข้อมูลนี้มาจากการรวม
กันของแหล่งเสียง 2 แหล่ง (sequence ดั้งเดิมของ
สัญญาณเสียง)

ปัญหา Cocktail Party: ตัวอย่างที่ 1



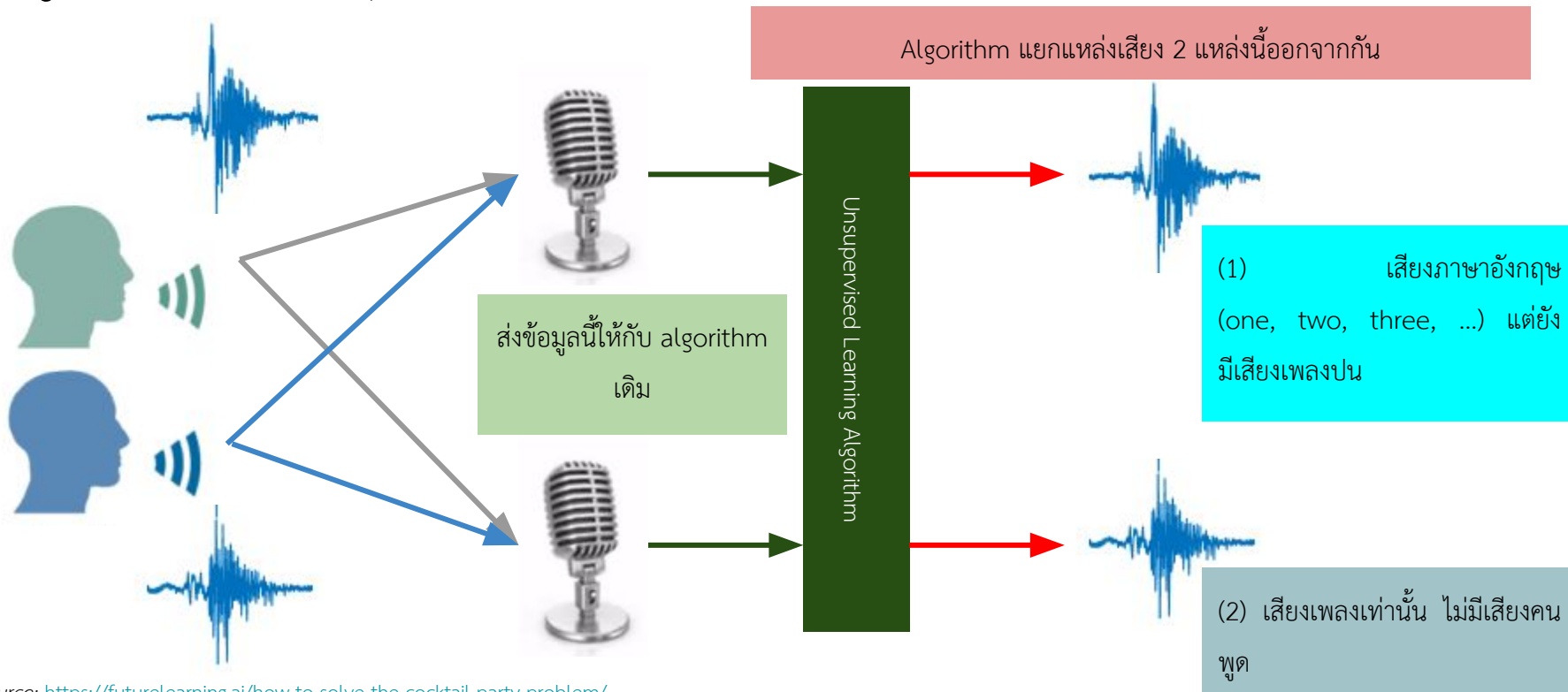
Source: <https://futurelearning.ai/how-to-solve-the-cocktail-party-problem/>

ปัญหา Cocktail Party: ตัวอย่างที่ 2



Source: <https://futurelearning.ai/how-to-solve-the-cocktail-party-problem/>

ปัญหา Cocktail Party: ตัวอย่างที่ 2



Source: <https://futurelearning.ai/how-to-solve-the-cocktail-party-problem/>

T

มันซับซ้อนขนาดไหนที่จะ implement algorithm นี้

เราทำ algorithm นี้ได้โดยเขียน code 1 บรรทัด

```
[W,s,v] = svd( ( repmat( sum(x.*x,1), size(x,1),1) .*x) *x' );
```

คำถามทบทวน

จากตัวอย่างต่อไปนี้ ตัวอย่างใดที่ควรใช้ UL algorithm แก่ (อาจมีข้อที่ถูกมากกว่า 1 ข้อ)

- (A) ให้ email ที่ถูกจัดว่าเป็น spam หรือ ไม่เป็น spam ให้ทำตัวกรอง email ที่เป็น spam (spam filter)
- (B) ให้บทความข่าวหลายๆอันที่พบจากเว็บ ให้จัดกลุ่มเป็นบทความที่เกี่ยวกับเรื่องเดียวกัน
- (C) ให้ฐานข้อมูลของข้อมูลลูกค้า หาส่วนตลาด (market segment) โดยอัตโนมัติ และจัดกลุ่มลูกค้าเป็น market segment ที่ต่างกัน
- (D) ให้ชุดข้อมูลคนไข้ที่ถูกวินิจฉัยว่าเป็นเบาหวานกับไม่เป็นเบาหวาน ให้เรียนรู้ที่จะแยกประเภท (classify) ว่าคนไข้คนใหม่ เป็นหรือไม่เป็นเบาหวาน

คำถามทบทวน

จากตัวอย่างต่อไปนี้ ตัวอย่างใดที่ควรใช้ UL algorithm แก่ (อาจมีข้อที่ถูกมากกว่า 1 ข้อ)

- (A) ให้ email ที่ถูกจัดว่าเป็น spam หรือ ไม่เป็น spam ให้ทำตัวกรอง email ที่เป็น spam (spam filter)
- (B) ให้บทความข่าวหลายๆอันที่พบจากเว็บ ให้จัดกลุ่มเป็นบทความที่เกี่ยวกับเรื่องเดียวกัน
- (C) ให้ฐานข้อมูลของข้อมูลลูกค้า หาส่วนตลาด (market segment) โดยอัตโนมัติ และจัดกลุ่มลูกค้าเป็น market segment ที่ต่างกัน
- (D) ให้ชุดข้อมูลคนไข้ที่ถูกวินิจฉัยว่าเป็นเบาหวานกับไม่เป็นเบาหวาน ให้เรียนรู้ที่จะแยกประเภท (classify) ว่าคนไข้คนใหม่ เป็นหรือไม่เป็นเบาหวาน

คำตอบ: (B) และ (C) → เพราะ (B) : เหมือนกับตัวอย่าง Google News ที่ใช้ clustering algorithm จัดกลุ่มบทความข่าว และ (C) : เป็นการแบ่งส่วนตลาด (market segmentation) ที่ใช้ algorithm หา market segment โดยอัตโนมัติ

สรุป: Unsupervised Learning

ทำให้เราสามารถ

- แก้ปัญหาโดยไม่ต้องรู้ว่าผลจะเป็นอย่างไร
- หาโครงสร้าง (structure) จากข้อมูลที่เราไม่รู้ผลของตัวแปรต่างๆ
- หาโครงสร้างได้โดยจัดกลุ่ม (cluster) ข้อมูล โดยใช้ความสัมพันธ์ระหว่างตัวแปรในข้อมูล

ไม่มี feedback เกี่ยวกับผลการทำนาย (prediction results)

ตัวอย่าง

Clustering: จากยีนที่ต่างกัน 1 ล้านกลุ่ม ให้หาวิธีจัดกลุ่มยีนที่คล้ายๆ กัน หรือมีตัวแปรที่เกี่ยวข้องกัน เช่น ช่วงชีวิต (lifespan), ตำแหน่ง, หน้าที่ เป็นต้น

Non-clustering: Cocktail Party Algorithm สามารถหาโครงสร้างในสิ่งแวดล้อมที่วุ่นวายได้ ก็คือ แยกเสียงของคนแต่ละคน และเสียงเพลงออกจากสัญญาณเสียงที่ปนรวมกัน

References

1. Andrew Ng, Machine Learning, Coursera.
2. Nessessence, อะไรคือ การเรียนรู้ของเครื่อง (Machine Learning)? (ฉบับมือใหม่), 18 Dec 2018, URL:
<https://www.thaiprogrammer.org/2018/12/%E0%B8%AD%E0%B8%B0%E0%B9%84%E0%B8%A3%E0%B8%84%E0%B8%B7%E0%B8%AD-%E0%B8%81%E0%B8%B2%E0%B8%A3%E0%B9%80%E0%B8%A3%E0%B8%B5%E0%B8%A2%E0%B8%99%E0%B8%A3%E0%B8%B9%E0%B9%89%E0%B8%82%E0%B8%AD%E0%B8%87%E0%B9%80/>
3. What is Machine Learning?, <https://www.digitalskill.org/contents/5>
4. คอมพิวเตอร์คลัสเตอร์, Computer Cluster คืออะไร, 11 Aug 2009, URL:
<https://www.ezylinux.com/%E0%B8%84%E0%B8%AD%E0%B8%A1%E0%B8%9E%E0%B8%B4%E0%B8%A7%E0%B9%80%E0%B8%95%E0%B8%AD%E0%B8%A3%E0%B9%8C%E0%B8%84%E0%B8%A5%E0%B8%B1%E0%B8%AA%E0%B9%80%E0%B8%95%E0%B8%AD%E0%B8%A3%E0%B9%8C-computer-cluster/>