ชื่อ_____ รหัสนักศึกษา _____

มหาวิทยาลัยสงขลานครินทร์

คณะวิศวกรรมศาสตร์

ภาควิชาวิศวกรรมคอมพิวเตอร์

สอบปลายภาค: ภาคการศึกษาที่ 1	ปีการศึกษา: 2556
วันที่สอบ: 11 ตุลาคม 2556	เวลาสอบ: 13.30 – 16.30
รหัสวิชา: 242 - 676	ห้องสอบ: S817
ชื่อวิชา: INTRO TO MACHINE LEARNING	

คำสั่ง: อ่านรายละเอียดของข้อสอบ และคำแนะนำให้เข้าใจก่อนเริ่มทำข้อสอบ

อนุญาต: เครื่องคิดเลขและเครื่องเขียนต่าง ๆ นำเข้าห้องสอบ

ไม่อนุญาต: เอกสารการสอน, อุปกรณ์สื่อสาร นำเข้าห้องสอบ

เวลา: 3 ชั่วโมง (180 นาที)

คำแนะนำ:

- ข้อสอบมี 11 หน้า (รวมใบปะหน้า) มี 10 ข้อ คิดเป็นคะแนนเก็บ 30 %
- คำตอบทั้งหมดจะต้องเขียนลงในข้อสอบ
- เขียนชื่อ รหัสนักศึกษา ในทุกหน้าของข้อสอบให้ชัดเจน

ทุจริตในการสอบ โทษขั้นต่ำคือ
ปรับตกในรายวิชาที่ทุจริต และพักการเรียน 1 ภาคการศึกษา

ชื่อ_____ รหัสนักศึกษา _____

1. Regression (3 คะแนน)

การหาพารามิเตอร์ของการเรียนรู้แบบ regression สำหรับชุดข้อมูลใด ๆ ที่สร้างจากสมการ polynomial กำลังต่าง ๆ สมมติให้ข้อมูลถูกสร้างจากสมการกำลัง 5 และมีค่าความผิดพลาดของ Gaussian ดังนี้

$$y = w_0 + w_1x + w_2x^2 + w_3x^3 + w_4x^4 + w_5x^5 + \varepsilon,\ \varepsilon \sim N(0,1)$$

ในการ training เรามี 100 ชุดข้อมูล {x,y} และอีก 100 ชุดข้อมูลสำหรับ testing และหากเราไม่รู้กำลังของ polynomial เราทำการเรียนรู้จากชุดข้อมูล โดยมีรูปแบบ 2 โมเดลคือ

โมเดลที่ 1 เรียนรู้พารามิเตอร์สำหรับ polynomial กำลัง 4

โมเดลที่ 2 เรียนรู้พารามิเตอร์สำหรับ polynomial กำลัง 6

จากทั้งสองโมเดลคิดว่าแบบใดจะเหมาะกับข้อมูลที่ใช้ในการ test มากที่สุดเพราะเหตุใด

ชื่อ_____ รหัสนักศึกษา _____

2. จงตอบคำถามต่อไปนี้ (3 คะแนน)

2.1 จงหาค่าของ VC dimension สำหรับ Linear Support Vector Machines ใน d มิติ

2.2 จากคำกล่าว "เส้นแบ่งการตัดสินใจ (decision boundary) ที่ได้จากโมเดลแบบ generative model ด้วยเงื่อนไขแบบ Gaussian distributions สามารถสร้างด้วยวิธีการ SVM โดยใช้ polynomial kernel" จงให้เหตุผลว่าจริงหรือเท็จอย่างไร

2.3 จากคำกล่าว "AdaBoost เป็นวิธีการเรียนรู้ที่สามารถทำให้การสอนไม่มีค่าผิดพลาด (zero training error) ได้โดยขึ้นอยู่กับชนิดของตัวแยกแยะที่ใช้ ว่ามีการนำเอาวิธีการแยกแยะแบบง่ายมารวมกันเพียงพอหรือไม่" จงให้เหตุผลว่าจริงหรือเท็จอย่างไร

2.4 เราสามารถหาค่าประมาณของจำนวนสถานะ (states) ของโมเดล Hidden Markov ได้อย่างไร

ชื่อ_____ รหัสนักศึกษา _____

3. Naïve Bayes (3 คะแนน)

3.1 กำหนดให้ X, Y และ Z เป็นตัวแปรแบบสุ่ม โดย $X \sim Unif(0,1)$ และ 0 < a < b < 1

$$Y = \begin{cases} 1 \; if \; 0 \leq X \leq a \\ 0 \qquad \text{อื่นๆ} \end{cases}$$

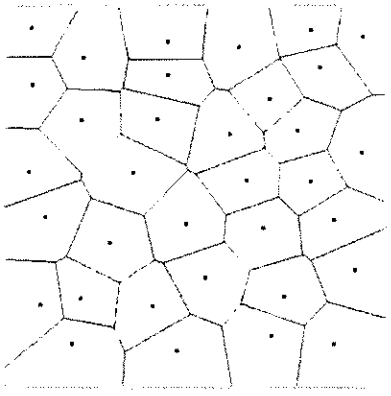
$$Z = \begin{cases} 1 \; if \; 0 \leq X \leq b \\ 0 \qquad \text{อื่นๆ} \end{cases}$$

จงแสดงให้เห็นว่า Yและ Z อิสระต่อกันหรือไม่ อย่างไร

ชื่อ_____ รหัสนักศึกษา _____

4. Nearest Neighbors (3 คะแนน)

กำหนดให้ข้อมูลต่อไปนี้



รูปที่ 1

จากรูปที่ 1 แสดงขอบเขตการตัดสินใจโดยวิธี 1 NN(Nearest Neighbors) โดยเป็นการแยกแยะข้อมูลในรูปแบบ 2 มิติบน Euclidean space กำหนดให้ n ตัวอย่างของจุด P1, P2, ..., Pn และกลุ่มข้อมูล(category/ class) ที่สัมพันธ์กันของแต่ละจุดคือ C1, C2, Cn (ค่าของ C ได้จากเซตของประเภทที่เป็นไปได้) วิธีการแยกแยะข้อมูลแบบ Nearest Neighbors จะพิจารณาข้อมูลใหม่ที่เข้ามาและแยกแยะตามเสียงส่วนมากที่อยู่ใน k ตัวอย่างใกล้เคียงด้วยการหาระยะแบบ Euclidean

4.1. หากต้องการสร้าง decision tree (โดยมีค่าแต่ละ node เป็น "is x > a", "is x < b", "is y > c", or "is y < d" กำหนดให้ a,b,c,d ค่าคงที่จำนวนจริง) ที่แยกแยะข้อมูลได้เช่นเดียวกับ 1-NN แบบใช้การวัดระยะของ Euclidean ได้หรือไม่ เพราะเหตุใด

4.2. สมมติให้ไม่ทราบวิธีการวัดระยะ แต่กำหนดกล่องดำหนึ่งขึ้นมา ซึ่งมีข้อมูลเข้าคือ ชุดของตัวอย่าง P1, P2, .. Pn และเมื่อข้อมูลใหม่ Q ผ่านกล่องดำนี้ ผลลัพธ์ที่ได้ของสมาชิกที่ใกล้สุด(nearest neighbor) ของ Q คือ Pi ซึ่งตรงกับ class ที่ชื่อ Ci ถามว่าเป็นไปได้หรือไม่ที่จะสร้างตัวแยกแยะแบบ k-NN จากกล่องดำอันนี้พร้อมให้เหตุผล

5. Decision Trees (3 คะแนน)

หากต้องการสร้าง Decision Tree สำหรับข้อมูลจำนวน n เวคเตอร์(vector)โดยแต่ละเวคเตอร์มี m แอตทริบิว (attribute)

5.1 กำหนดให้ i และ j เป็น attribute ของ training data ในเวคเตอร์ X โดยค่าของ $x_i = x_j$ (x_i คือค่าลำดับที่เข้ามาในเวคเตอร์ X) ในทุก ๆ ตัวอย่าง สมมติเราเลือกใช้ x_i ตัวเดียว เนื่องจาก i และ j เป็นข้อมูลเดียวกัน เราสามารถเอาค่า x_j ออกจาก training data โดยไม่ทำให้ decision tree เปลี่ยนใช่หรือไม่ เพราะเหตุใด

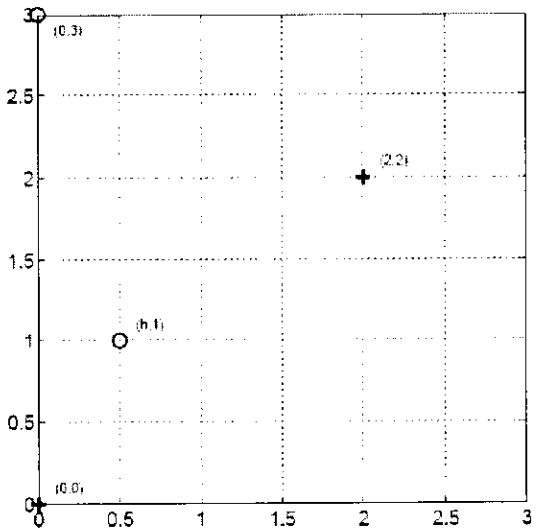
5.2 กำหนดให้ X และ Z เป็น 2 เวคเตอร์ที่เท่ากันในชุดข้อมูลของการสอน (training set) โดยค่าต่าง ๆ ของ X และ Z ไม่ว่าจะเป็น attributes รวมถึงกลุ่ม (label) มีค่าเท่ากันหมด หากว่าเราลบ Z ออกจาก training data แล้วจะทำให้ decision tree เปลี่ยนหรือไม่ จงอธิบาย

ชื่อ_____ รหัสนักศึกษา _____

5.3 สำหรับการตัดสินใจแบบ 3 ทาง โดยกำหนดพารามิเตอร์จาก {a,b} สามารถนำอาการตัดสินใจแบบ 2 ทางมาใช้เพื่อทำการตัดสินใจให้ได้ผลลัพธ์แบบเดียวกันได้หรือไม่ จงอธิบายโดยการวาดภาพ

6. Support Vector Machine (3 คะแนน)

กำหนดให้มีข้อมูลในการสอน (training data) เพียง 4 ข้อมูลใน 2 มิติ ดังรูปที่ 2



รูปที่ 2

กำหนดให้ positive examples คือ x1 = [0, 0] , x2 = [2, 2] และ negative examples คือ x3 = [h, 1] , x4 = [0, 3], โดยให้ $0 \leq h \leq 3$ เป็นค่าพารามิเตอร์

6.1 ถ้าค่าของ $h \geq 0$ และมีค่ามากสุดเป็นเท่าไหร โดยที่ training points ยังเป็น linearly separable อยู่

6.2 ค่าของ maximum margin จะมีการเปลี่ยนแปลงตามค่า h หรือไม่

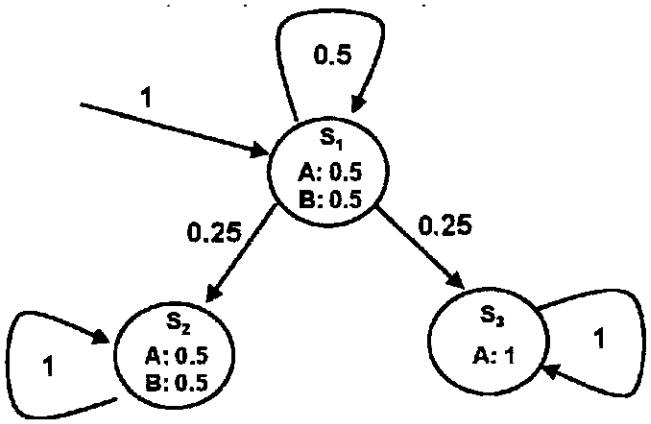
ชื่อ_____ รหัสนักศึกษา _____

6.3 จงหาค่า maximum margin boundary ตามฟังก์ชันของ h

[ข้อสังเกต : เป็นสมการเส้นตรงโดยเมื่อ h = 1 ค่า maximum margin เป็น 0]

7. HMM (3 คะแนน)

กำหนดให้ HMM states ดังรูปที่ 3



รูปที่ 3

สิ่งที่สนใจคือ observed outputs โดยจะอยู่ในรูปของ $P(O_1 \dots O_T)$ หากกำหนดให้

$v_i^t = P(O_1 \dots O_T | q_t = s_i)$

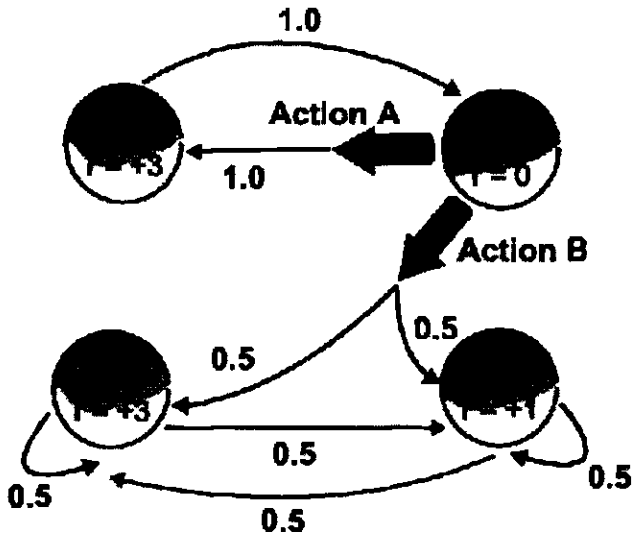
7.1 จงหาค่า $P(O_1 \dots O_T)$ ในรูปของ v_i^t และ $p_t(i)$ โดยกำหนดให้ $p_t(i) = \mathrm{p}(q_t = s_i)$

ชื่อ_____ รหัสนักศึกษา _____

7.2 จงหาค่า $P(O_1 = B \dots O_{200} = B)$ (เกิด B อยู่ในแถว 200 ครั้ง) ในรูปของ v_i^t และ $p_t(i)$ โดยกำหนดให้ $p_t(s_2) = p_t(s_3)$ แนะนำให้กำหนดค่า t ที่เหมาะสมก่อนแล้วจากนั้นนำค่าดังกล่าวไปใช้คำนวนค่าผลลัพธ์

8. Markov Decision Process (3 คะแนน)

กำหนดให้ Markov Decision Process ดังรูปที่ 4 โดยให้ r คือ ค่ารางวัล (reward) ที่ได้ในแต่ละ state



รูปที่ 4

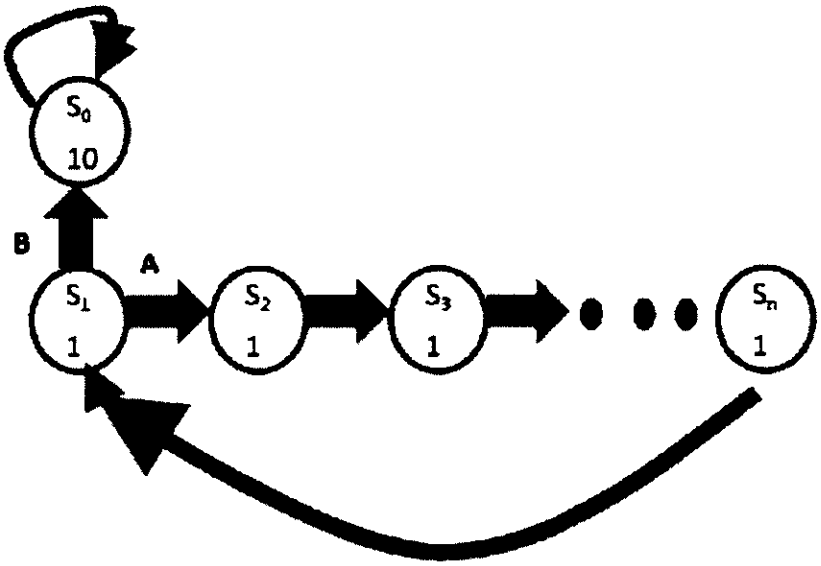
8.1 จากรูปที่ 4 Action A หรือ B ที่ให้ค่า reward สูงสุดหากให้จุดเริ่มต้นที่ State 2

8.2 กำหนดให้ γ เป็นค่า discount factor เป็น 0.9 จงหาค่า discounted future reward ของทั้งสอง Action และ Action ใดที่ให้ค่ารวมสูงสุดของ discounted future reward

ชื่อ_____ รหัสนักศึกษา _____

9. Reinforcement Learning (3 คะแนน)

กำหนดให้ รูปที่ 5 คือโมเดลของ Reinforcement Learning โดยให้ค่า n > 1000 และ ทุก state มี action ที่เป็นไปได้(deterministic outcome) เพียง 1 ยกเว้น state ที่ 1 มี action เป็นไปได้ 2 คือ A และ B โดย action A จะนำไปสู่ S2 และ Action B จะนำไปสู่ S0 สมมติให้ ค่า γ (discount factor) = 0.5



รูปที่ 5

เพื่อเรียนรู้โมเดลนี้ เราจะใช้ Q learning ด้วยค่า $\alpha = 1$ โดยค่าฟังก์ชัน Q สำหรับทุก states มีค่าเริ่มต้นเป็น 0 เมื่อมาถึง S1 จะใช้ฟังก์ชัน Q ในการประมาณค่า Action ต่อไปจากค่าสูงสุด และเริ่มต้นจาก State S1

9.1 หลังจาก 1 step จงหาค่า Q(S1, A) และ Q(S1, B)

_____

_____

_____

9.2 หลังจาก 5 step จงหาค่า Q(S1, A) และ Q(S1, B)

_____

_____

_____

9.3 หลังจาก n+5 step จงหาค่า Q(S1, A) และ Q(S1, B)

_____

_____

_____

ชื่อ_____ รหัสนักศึกษา _____

10. Dimensionality Reduction (3 คะแนน)

กำหนดให้ ตารางที่ 1 แสดงข้อมูล โดยกำหนดให้ ค่าส่วนประกอบหลักแรก(first principle component) คือ (0.694, 0.720)

ข้อมูล	x	y
1	5.51	5.35
2	20.82	24.03
3	-0.77	-0.57
4	19.30	19.38
5	14.24	12.77
6	9.74	9.68
7	11.59	12.06
8	-6.08	-5.22

ตารางที่ 1

10.1 จงหาค่าที่แสดงแทน (projected coordinate) ค่าข้อมูลชุดที่ 1 (5.51,5.35)

10.2 จงหาค่าตำแหน่งของ x,y ข้อมูลชุดที่ 1 (5.51,5.35) ที่ได้จากการสร้างค่าใหม่(reconstruct) โดยใช้ first principle component

10.3 จงหาค่าที่แสดงแทน (projected coordinate) ค่าข้อมูลชุดที่ 1 (5.51,5.35) ใน second principle component space

10.3 การสร้างค่าใหม่ (reconstruct) ค่าเริ่มต้นของข้อมูลโดยใช้ second principle component จะมีค่าผิดพลาดเป็นเท่าไร