

การบ้านที่ 1 (1/2564)

รายวิชา คพ.348 แบบจำลองสำหรับวิทยาการข้อมูล

กำหนดส่ง วันที่ 2 ตุลาคม 2564 (ก่อนเวลา 23:55 น.)

คำชี้แจง/ข้อกำหนด

1. การบ้านชิ้นนี้เป็นการบ้านรายบุคคล นักศึกษาจะต้องทำโจทย์ปัญหาด้วยตนเองทุกข้อ หากมีการสงสัยว่านักศึกษามีการทุจริตไม่ว่าด้วยวิธีการใด ผู้สอนรายวิชานี้จะเรียกสอบสวน หากพบว่าการทุจริตจริง นักศึกษาจะไม่ได้รับคะแนนของการบ้านชิ้นนี้ และผู้สอนขอสงวนสิทธิ์ในการดำเนินการตามระเบียบมหาวิทยาลัยจนถึงที่สุด

2. การบ้านมีทั้งหมด 4 ข้อ คะแนนเต็ม 100 คะแนน ดังนี้

ข้อที่	1	2	3	4
คะแนนเต็ม	25	25	30	20

3. การส่งการบ้านล่าช้าเกิน 5 วัน หลังจากกำหนดส่ง ผู้สอนขอสงวนสิทธิ์ในการไม่ตรวจให้คะแนนการบ้านไม่ว่ากรณีใดทั้งสิ้น แต่หากหากนักศึกษาส่งการบ้านล่าช้าภายใน 5 วัน คะแนนของการบ้านชิ้นนี้จะถูกหักร้อยละ 20 ของคะแนนเต็มต่อ 1 วันที่ส่งล่าช้า (โดยนักศึกษามีสิทธิในการขอละเว้นการตัดคะแนนในกรณีส่งล่าช้านี้ 1 ครั้งจากจำนวนการบ้านทั้งหมด)
4. ให้นักศึกษาส่งการบ้านในส่วนที่เป็นข้อเขียน (ถ้ามี) ในรูปแบบไฟล์ PDF โดยมีรูปแบบการตั้งชื่อไฟล์ดังนี้ CS348_XXXXXXXXXX_1-2564.pdf โดยที่ XXXXXXXXXXXX คือรหัสนักศึกษา ให้นักศึกษาระบุข้อที่ตอบให้ชัดเจนสำหรับคำตอบของปัญหาแต่ละข้อ ลงในไฟล์เดียว
5. ให้นักศึกษาส่งการบ้านในส่วนที่เป็นการบ้านเขียนโปรแกรม (ถ้ามี) ในรูปแบบไฟล์ของภาษาโปรแกรมนั้น ๆ โดยมีรูปแบบการตั้งชื่อไฟล์ดังนี้ CS348_XXXXXXXXXX_YY_1-2564.ZZZ โดยที่ XXXXXXXXXXXX คือรหัสนักศึกษา 10 หลัก YY คือข้อที่นักศึกษาเขียนโปรแกรมเพื่อแก้ปัญหา และ ZZZ คือ file extension ที่สอดคล้องกับภาษาโปรแกรมที่นักศึกษาใช้ ตัวอย่าง ถ้านักศึกษาใช้ภาษา C++ เพื่อแก้ปัญหาข้อ 3 ให้นักศึกษาดังชื่อไฟล์ดังนี้ CS348_6209610123_03_1-2564.cpp
6. **นักศึกษาไม่ต้องส่ง Hard copy ของการบ้านฉบับนี้ เนื่องจากสถานการณ์ COVID-19**
7. หากการส่งการบ้านไม่ตรงตามข้อกำหนดในข้อ 4., 5. และ 6. การบ้านชิ้นนี้อาจไม่ได้รับการตรวจให้คะแนน
8. นักศึกษามีสิทธิในการโต้แย้งคะแนนของนักศึกษาสำหรับการบ้านชิ้นนี้ภายใน 7 วันหลังจากการประกาศคะแนน หลังจากนั้นจะถือว่านักศึกษายอมรับคะแนนการบ้านชิ้นนี้โดยปราศจากข้อโต้แย้งใด ๆ

1. (ข้อเขียน, 25 คะแนน)

- 1.1. ให้นักศึกษาอธิบายความหมายของ Data Science ตามความเข้าใจของนักศึกษา (5 คะแนน)
- 1.2. ให้นักศึกษาเข้าไปที่เว็บไซต์ <https://data.go.th/> (หรือแหล่งข้อมูลอื่นที่นักศึกษาสงใจ) (20 คะแนน)
 - 1.1.1 เลือกข้อมูลที่นักศึกษาสงใจมา 3 แบบ
 - 1.1.2 อธิบายรายละเอียดเกี่ยวกับข้อมูลทีเลือกมาแต่ละแบบ และระบุ meta data
 - 1.1.3 ตั้งคำถามที่น่าสนใจเกี่ยวกับข้อมูลแต่ละแบบทีเลือกมา

2. (ข้อเขียน, 25 คะแนน)

- 2.1. ให้นักศึกษาบอกความหมายของ แบบจำลองความน่าจะเป็น (probabilistic model) ตามความเข้าใจของนักศึกษา (5 คะแนน)
- 2.2. ให้นักศึกษาแสดงวิธีการหาคำตอบต่อไปนี้ (20 คะแนน)
 - 2.2.1. นักศึกษาเริ่มทอยลูกเต๋า 4 หน้า (แต้มที่ออกได้คือ 1-4) 1 ลูกได้แต้ม N หลังจากการทอยลูกเต๋า นักศึกษาโยนเหรียญ 2 หน้า (H/T) 1 เหรียญ จำนวน N ครั้ง ให้เขียน Sample space ของการทดลองนี้
 - 2.2.2. ลูกเต๋า 6 หน้า (แต้มที่ออกได้คือ 1 ถึง 6) โดยมีความน่าจะเป็นของการออกหน้าลูกเต๋าดั้มคู่เป็นสองเท่าของลูกเต๋าดั้มคี่ และ ความน่าจะเป็นทีลูกเต๋าคจะออกดั้มคู่มีค่าเท่ากัน (รวมถึงการออกดั้มคี่ด้วย) ให้นักศึกษาสร้างแบบจำลองความน่าจะเป็นการทดลองโยนลูกเต๋าลูกนี้ 1 ครั้ง และ หาคความน่าจะเป็นทีลูกเต๋าคจะออกดั้มน้อยกว่า 4
 - 2.2.3. จากการโยนเหรียญ 2 หน้า ปราศจากอคติ 1 เหรียญ 2 ครั้ง ให้ A แทนเหตุการณ์ทีการโยนครั้แรกออกหัว และ B แทนเหตุการณ์ทีผลลัพธ์การโยนเหรียญทั้งสองครั้ออกหน้าผลลัพธ์เดียวกัน ให้นักศึกษาแสดงว่า A และ B เป็นเหตุการณ์ทีเป็นอิสระต่อกัน (Independent) หรือไม
 - 2.2.4. ในพื้นที่หนึ่ง ถ้าเครื่องบินรบปรากฏขึ้นเรดาร์จะสามารถแจ้งเตือนได้ด้วยความน่าจะเป็น 0.99 ในขณะที่มีความน่าจะเป็นที 0.10 ทีเรดาร์นี้จะแจ้งเตือนโดนทีไม่มีเครื่องบินรบปรากฏขึ้น (False alarm) ถ้าสมมติว่าเรารู้ว่าเหตุการณ์ทีเครื่องบินรบจะปรากฏขึ้นอยู่ทีความน่าจะเป็น 0.05 หาคความน่าจะเป็นทีเครื่องบินรบจะปรากฏขึ้นโดยทีเรดาร์มีการแจ้งเตือน

3. (ข้อเขียน, 30 คะแนน)

- 3.1. สมมติว่านักศึกษาไปงานเลี้ยงแห่งหนึ่งทีมีคนร่วมงานทั้งหมด 500 คน จงหาความน่าจะเป็นทีคนในงานคนอื่นจะมีวันเกิดวันเดียวกับคุณ (สมมติว่า 1 ปีมี 365 วัน) (Hint: Binomial, Poisson) (5 คะแนน)
- 3.2. กำหนดให้ X เป็นตัวแปรสุ่มทีรับค่าจำนวนเต็มตั้งแต่ 0 ถึง 9 ด้วยความน่าจะเป็นทีเท่ากันคือ $\frac{1}{10}$ (10 คะแนน)
 - 3.2.1. หา PMF ของตัวแปรสุ่ม $Y = X \bmod (3)$
 - 3.2.2. หา PMF ของตัวแปรสุ่ม $Y = 5 \bmod (X + 1)$
- 3.3. ให้ a และ b เป็นจำนวนเต็ม โดย $a \leq b$ ให้ X เป็นตัวแปรสุ่มทีรับค่าในช่วง $[2^a, 2^b]$ ด้วยความน่าจะเป็นทีเท่ากันโดยค่าที X รับผิดชอบอยู่ในรูปของ 2^i โดย i เป็นจำนวนเต็มทีอยู่ในช่วง $a \leq i \leq b$ ให้นักศึกษาหา Expected value และ Variance ของ X (10 คะแนน)

3.4. ให้นักศึกษาแสดงว่า ถ้า ตัวแปรสุ่ม $Y = aX + b$ โดย a และ b เป็น scalars แล้ว $var(Y) = a^2 var(X)$ (5 คะแนน)

4. (เขียนโปรแกรม-รายงาน, 20 คะแนน)

Quiz 05