

รายละเอียดของกระบวนวิชา 229351 ปีการศึกษา 2/2564

1. รหัสกระบวนวิชา 229351
ชื่อกระบวนวิชา การเรียนรู้เชิงสถิติสำหรับวิทยาการข้อมูล (Statistical Learning for Data Science 1)
2. ผู้สอน อาจารย์ ดร. ตลภาค พรนพรัตน์
3. เวลา Sec 01 : Lecture M-Th 11:00-13:00 น. ใน Zoom (Link อยู่ใน Microsoft Teams)
4. หน่วยกิต 3 (3-0-6)

แผนการสอน

สัปดาห์ที่	หัวข้อ/รายละเอียด	จำนวนชั่วโมง			กิจกรรมการเรียนการสอน
		บรรยาย	ปฏิบัติ	ฝึกปฏิบัติ	
1	ชี้แจงและแนะนำกระบวนวิชา แนวคิดพื้นฐานของวิทยาการข้อมูล - ทบทวนหลักการเขียนภาษา python	2	2	0	การบรรยายโดยใช้ภาพนิ่งและคอมพิวเตอร์
2-3	พีชคณิตเชิงเส้นสำหรับวิทยาการข้อมูล (Linear algebra) - นิยามและคุณสมบัติของเวกเตอร์และเมทริกซ์ - การคำนวณด้วยเวกเตอร์และเมทริกซ์ - การแปลงเวกเตอร์ด้วยเมทริกซ์ - การแยกส่วนประกอบของเมทริกซ์	2	2	0	- การบรรยายโดยใช้ภาพนิ่งและคอมพิวเตอร์ - ปฏิบัติการคอมพิวเตอร์ด้วยภาษา python
4	การวิเคราะห์องค์ประกอบหลัก (Principal component analysis)	4	4	0	
5	- นิยามของการเรียนรู้เชิงสถิติ - การสร้าง Training set และ Test set - Bias-Variance tradeoff, no-free-lunch theorem - การวัดประสิทธิภาพของโมเดล	2	2	0	
6-8	การวิเคราะห์การถดถอย (Linear regression) - การเตรียมข้อมูลเพื่อการวิเคราะห์การถดถอย ➢ การสร้างตัวแปรและตัวแปรหุ่น ➢ การสร้างเมทริกซ์แผนแบบ - การวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย - การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ - การทดสอบความสัมพันธ์ของตัวแปร - การเลือกตัวแปร - การถดถอยแบบบริดจ์และแบบ Lasso	8	8	0	
9-12	การวิเคราะห์อนุกรมเวลาสำหรับวิทยาการข้อมูล - การเตรียมข้อมูลเพื่อการวิเคราะห์อนุกรมเวลา - การวิเคราะห์อนุกรมเวลาแบบคลาสสิก ➢ ความเป็นฤดูกาลและความคงที่ ➢ สหสัมพันธ์ในตัวเองและสหสัมพันธ์ในตัวเองบางส่วน ➢ เทคนิคการปรับให้เรียบ ตัวแบบการถดถอยในตัว ตัวแบบการเคลื่อนที่ในตัว ตัวแบบการเคลื่อนที่ที่ถดถอยในตัว	8	8	0	

13-15	การวิเคราะห์การถดถอยโลจิสติกส์สำหรับวิทยาการข้อมูล – การเตรียมข้อมูลเพื่อการวิเคราะห์การถดถอยโลจิสติกส์ทวิ – การวิเคราะห์การถดถอยโลจิสติกส์ทวิ ➤ วิธีภาวน่าจะเป็นสูงสุด – การวิเคราะห์การถดถอยโลจิสติกส์ทุกกลุ่ม – การวิเคราะห์การถดถอยโลจิสติกส์อันดับ	8	8	0	
-------	--	---	---	---	--

การประเมินผลการเรียนรู้

2.1 สรุปสัดส่วนการให้คะแนน

- ปฏิบัติการ 12 ครั้ง คะแนน 20%
- Kaggle คะแนน 20%
- สอบกลางภาค คะแนน 30% (TBD)
- สอบปลายภาค คะแนน 30% (สอบวันพุธ ที่ 2 มีนาคม 2565 เวลา 12:00-15:00 น.)

2.2 การประเมินผลการเรียนการสอน

- นักศึกษาต้องเข้าสอบทั้งกลางภาคและปลายภาค หากนักศึกษาขาดสอบอย่างใดอย่างหนึ่ง นักศึกษาจะได้รับลำดับชั้น F
- วิธีการตัดเกรดแบบอิงกลุ่ม

ปฏิบัติการ (Lab)

- ทุก ๆ Thursday
- เน้นการวิเคราะห์ข้อมูลด้วยภาษา Python เราจะทำปฏิบัติการผ่าน Google Colab
- Download หรือ Print Notebook เป็นไฟล์ pdf แล้วส่งไฟล์ใน Microsoft Teams

Kaggle

- สมัครเป็นสมาชิกเว็บไซต์ kaggle แล้วเข้าไปที่เว็บไซต์ของการแข่งขัน (link อยู่ใน Microsoft Teams)
- สร้างตัวแบบในการทำนายที่ดีที่สุดโดยใช้ training set แล้ว upload การทำนายบน test set
- กลุ่มละ 1-3 คน การแข่งขันสิ้นสุดวันพุธที่ 23 กุมภาพันธ์ 2565 เวลา 23:59 น.
- เขียนรายงานถึงวิธีที่ใช้และประสิทธิภาพของตัวแบบที่ได้ ความยาว 3-15 หน้า ส่งภายในวันพุธ ที่ 2 มีนาคม 2565

ทรัพยากรประกอบการเรียนการสอน

1 ตำราและเอกสารหลัก เอกสารประกอบการสอนของอาจารย์
2 เอกสารและข้อมูลแนะนำ <ol style="list-style-type: none"> 1. Trevor Hastie et al. An Introduction to Statistical Learning: with Applications in R. (available online at https://faculty.marshall.usc.edu/gareth-james/ISL/ISLR Seventh Printing.pdf) python code of examples in this book: https://github.com/JWarmerhoven/ISLR-python) 2. Rob J Hyndman and George Athanasopoulos, Forecasting: Principles and Practice. (available online at https://otexts.com/fpp2) 3. Gilbert Strang, Linear Algebra and Learning from Data.