

האוניברסיטה העברית בירושלים

סילבוס

למידה סטטיסטית למדע הנתונים - 52016

תאריך עדכון אחרון 20-12-2023

נקודות זכות באוניברסיטה העברית: 4

תואר: מוסמך

היחידה האקדמית שאחראית על הקורס: סטטיסטיקה

השנה הראשונה בתואר בה ניתן ללמוד את הקורס: 0

סמסטר: סמסטר ב'

שפת ההוראה: עברית

קמפוס: הר הצופים

מורה אחראי על הקורס (רכז): ברק סובר

דוא"ל של המורה האחראי על הקורס: Barak.Sober@mail.huji.ac.il

שעות קבלה של רכז הקורס:

מורי הקורס:

ד"ר ברק סובר,
מר גדעון יופה

תאור כללי של הקורס:

הקורס עוסק בשיטות ניתוח נתונים מודרניים בסטטיסטיקה, בעיקר בנתונים בהיקף גדול וממימד גבוה. נדון באתגרים סטטיסטיים וחשובים העולים מנתונים אלו, כולל עקרונות וכן שיטות פרקטיות. חלק מרכזי מהקורס גלום במעבדות (עבודות בית) בהן הסטודנטים מנתחים נתונים אמיתיים וחוקרים שיטות בעזרת סימולציות.

מטרות הקורס:

מטרת הקורס היא להציג בפני הסטודנטים שיטות מודרניות לניתוח נתונים סטטיסטיים. כמו כן, המטרה היא שסטודנטים יתנסו בעקרונות של עבודה מחקרית הן בתחום ניתוח הנתונים והן בתחום השוואה של שיטות אמידה / ניבוי.

תוצרי למידה

בסיומו של קורס זה, סטודנטים יהיו מסוגלים:

בסיומו של קורס זה, סטודנטים יהיו מסוגלים:

- לבחון קובץ נתונים ולהציג אותו.

- לנסח שאלה מחקרית כבעיית חיזוי, ולהבין את הייתרונות והחסרונות של חיזוי (לעומת סוגים שונים של הסקה סטטיסטית)

- לבנות מודל חיזוי (קטגוריאלי או רציף)

- לכמת את הצלחת המודל ולהשוות בין מודלים שונים ובין שיטות שונות. לאמוד את השגיאה ואת חוסר הוודאות.

- להציג בכתב את מסקנות הניתוח

דרישות נוכחות (%) :

80

שיטת ההוראה בקורס: הקורס מעובר פרונטאלית בכיתה. הנוכחות חובה. שיעורים בודדים יועברו בזום (הודעה תנתן מראש). השיעורים יוקלטו אך תנתן גישה רק לתלמידים אשר יראו סיבה לגיטימית להעדרות.

רשימת נושאים / תכנית הלימודים בקורס:

הרשימה מטה איננה רשימה מלאה או מחייבת. ייתכן שחלק מהנושאים ירדו ושאחרים יצטרפו. זוהי רשימה טנטטיבית.

1. ניקוי והכרת קובץ נתונים

2. PCA

-
3. ייצוג ומרחקים
 4. Clustering
 5. ניתוח יציבות ו Bootstrap
 6. מבוא ללמידה מבוקרת (learning Supervised). הטיה לעומת שונות.
 7. רגרסיה מתקדמת: הרחבת בסיסי
 8. רגולריזציה
 9. עצי רגרסיה
 10. קלאסיפיקציה: מודלים גנרטיביים
 11. SVM ו Reproducing Kernels
 12. Boosting
 13. מבוא לרשתות נוירונים

חומר חובה לקריאה:
אין

חומר לקריאה נוספת:
Elements of Statistical Learning (Hastie, Tibshirani & Friedman)
<https://web.stanford.edu/~hastie/ElemStatLearn/>

Understanding Machine Learning (Shalev-Shwartz & Ben-David)
<https://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/understanding-machine-learning-theory-algorithms.pdf>

A Distribution-Free Theory of Nonparametric Regression (Gyorfi, Kohler, Krzyzak & Walk)
<https://web.stanford.edu/class/ee378a/books/book1.pdf>

מרכיבי הציון הסופי :
הגשת עבודה מסכמת / פרויקט גמר / מטלת סיכום / מבחן בית / רפרט 75 %
מטלות הגשה במהלך הסמסטר: תרגילים / עבודות / מבדקים / דוחות / פורום / סימולציה ואחרות
25 %

מידע נוסף / הערות: