



סילבוס

## למידה סטטיסטית וניתוח נתונים - 52525

תאריך עדכון אחרון 04-03-2019

נקודות זכות באוניברסיטה העברית: 4

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): יובל בנימיני

דוא"ל של המורה האחראי על הקורס: [yuval.benjamini@mail.huji.ac.il](mailto:yuval.benjamini@mail.huji.ac.il)

שעות קבלה של רכז הקורס:

מורי הקורס:

ד"ר יובל בנימיני  
מר בועז רפאל מתן

תאור כללי של הקורס:

---

הקורס עוסק בשיטות ניתוח נתונים מודרניים בסטטיסטיקה, בעיקר בנתונים בהיקף גדול וממימד גבוה. נדון באתגרים סטטיסטיים וחשוביים העולים מנתונים אלו, כולל עקרונות וכן שיטות פרקטיות. חלק מרכזי מהקורס גלום במעבדות (עבודות בית) בהן הסטודנטים מנתחים נתונים אמתיים (כולל דפוסי הצבעה, התבטאות גניים, ותגובה מוחית), 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 חיזויי (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 (%):

0

שיטת ההוראה בקורס: הרצאה ותרגול.

רשימת נושאים / תכנית הלימודים בקורס:

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. מודלים דמויי רגרסיה + SVM
12. Boosting

---

חומר חובה לקריאה:  
אין

חומר לקריאה נוספת:  
*Advanced Data Analysis from an Elementary Point of View, Cosma Rohilla Shalizi*  
<http://www.stat.cmu.edu/~cshalizi/ADAfaEPoV/>

*The Elements of Statistical Learning* □ *Data mining, inference and prediction*  
(Tibshirani, Hastie and Friedman)  
<http://www-stat.stanford.edu/~tibs/ElemStatLearn/>

הערכת הקורס - הרכב הציון הסופי :  
מבחן מסכם בכתב/בחינה בעל פה 0 %  
הרצאה 0 %  
השתתפות 0 %  
הגשת עבודה 24 %  
הגשת תרגילים 76 %  
הגשת דו"חות 0 %  
פרויקט מחקר 0 %  
בחנים 0 %  
אחר 0 %

מידע נוסף / הערות: