

האוניברסיטה העברית בירושלים

סילבוס

עיבוד נתונים בסטטיסטיקה מודרנית - 52311

תאריך עדכון אחרון 07-05-2024

נקודות זכות באוניברסיטה העברית: 4

תואר: בוגר

היחידה האקדמית שאחראית על הקורס: סטטיסטיקה

השנה הראשונה בתואר בה ניתן ללמוד את הקורס: 2024

סמסטר: סמסטר ב'

שפת ההוראה: עברית

קמפוס: קרית א"י ספרא

מורה אחראי על הקורס (רכז): אריאל יפה

דוא"ל של המורה האחראי על הקורס: ariel.jaffe@mail.huji.ac.il

שעות קבלה של רכז הקורס: חמישי 12-13

מורי הקורס:

תאור כללי של הקורס:

הקורס יעסוק בשיטות ניתוח נתונים בהיקף גדול וממימד גבוה. נדון באתגרים חישוביים וסטטיסטיים העולים מניתוח נתונים אלו, בדגש על שיטות פרקטיות ויעילות חישובית. במהלך הקורס נלמד ונממש שיטות סטטיסטיות שונות, ונשתמש בהן לניתוח נתונים הן נתונים מסומלצים והן נתונים אמיתיים מתחומים שונים.

מטרות הקורס:

ממטרת הקורס היא להציג לסטודנטים שיטות מודרניות בסטטיסטיקה וניתוח נתונים

תוצרי למידה

בסיומו של קורס זה, סטודנטים יהיו מסוגלים:

להבין מספר שיטות סטטיסטיות מודרניות, לממש אותם בשפת תכנות סטנדרטית באופן יעיל, ולהפעיל אותן על נתונים אמפיריים על מנת לפתור בעיה מדעית קונקרטית.

דרישות נוכחות (%) :

חמישים אחוז נוכחות חובה

שיטת ההוראה בקורס: הרצאה ותרגול.

רשימת נושאים / תכנית הלימודים בקורס:

רשימה התכנים יכולה להשתנות:

0. עיבוד מקדים של נתונים: נורמליזציה, השלמת מידע חסר, ויזואליזציה
1. בדיקת השערות: חישובי עוצמה, חישוב $value-p$ באמצעות סימולציות, שיטות לבדיקת השערות מרובות
2. שיטות רגרסיה: רגרסיה לינארית מרובת משתנים, שיטות לבחירת משתנים וייצוג דליל
3. שיטות קלסיפיקציה: רגרסיה לוגיסטית, CART, רשתות נוירונים
6. שיטות לבחירת מודל ומיצוע על מודלים שונים
5. שיטות לינאריות ולא לינאריות להורדת מימד:
- PCA, manifold learning
6. אלגוריתמים למציאת צברים

חומר חובה לקריאה:

חומר לקריאה נוספת:

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

Large Scale Inference, Bradley Efron
http://statweb.stanford.edu/~ckirby/brad/LSI/monograph_CUP.pdf

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

מרכיבי הציון הסופי:

הגשת עבודה מסכמת / פרויקט גמר / מטלת סיכום / מבחן בית / רפרט 50 %
מטלות הגשה במהלך הסמסטר: תרגילים / עבודות / מבדקים / דוחות / פורום / סימולציה ואחרות
45 %
נוכחות / השתתפות בסיוור 5 %

מידע נוסף / הערות:
(יעודכן בהמשך)