

תאריך עדכון: 25/06/18

## שיטות במדעי הנתונים לפיסיקאים

86-605-01

ד"ר אמיר בשן

סוג הקורס: הרצאה

שנת לימודים: תשע"ט סמסטר: ב' היקף שעות: 3

אתר הקורס באינטרנט: [/http://physics.biu.ac.il](http://physics.biu.ac.il)

א. מטרת הקורס ותוצרי למידה (מטרות על / מטרות ספציפיות):  
היכרות תיאורטית ומעשית (hands-on) עם מגוון נושאים במדעי הנתונים.

תוצרי למידה:

1. להיחשף ולהכיר מושגים בסיסיים במדעי הנתונים.
2. להכיר את הבסיס התיאורטי והעקרונות היסודיים של שיטות שימושיות במדעי הנתונים.
3. להבין את היתרונות והחסרונות בשיטות השונות.
4. להתנסות באופן מעשי בנייתוה נתונים אמתיים תוך הבנת האתגרים המיוחדים הנובעים מהם.

ב. תוכן הקורס:

רציונל, נושאים:

בכל שבוע נעסוק בנושא אחר במדעי הנתונים. ההרצאה תעסוק בעקרונות הבסיסיים והתרגול יעסוק בהדרכה מעשית. על כל נושא יינתן תרגיל בית שבו הסטודנטים יידרשו ליישם את החומר הנלמד על נתונים אמתיים. נעסוק בענפים העיקריים ששייכים למדעי הנתונים, סטטיסטיקה ולמידת מכונה, תוך מתן דגש על נתונים המתקבלים ממערכות פיסיקליות.

## תכנית הוראה מפורטת לכל השיעורים:

All exercises will focus on real or simulated data. For each data type there should be some preprocessing of noise reduction if necessary.

1. Probability density functions, Mean, Variance, Bessel's Correction, Correlations, Correlation types, Entropy, Simulating Data.
2. Likelihood, Maximum Likelihood Estimation, Chi-square test, statistical significance, p-value, null-hypothesis, Bootstrapping and other resampling analysis methods.
3. Linear Regression, Binary Classifier - Logistic Regression, SVM, Overfitting and Underfitting, False/True Negative/Positive - Confusion Matrix, Receiver Operating Characteristic,  $R^2$ , Cross-Validation and other goodness-of-fit measures.
4. Perceptron, Neural Network Classifier, Neural Network Fitting, Bias in Neural Networks, real data fitting, Digits classification, comparison with other methods.
5. Convolutional Neural Network.
6. Clustering algorithms, k-means, tsne, PCA, PCOA, Factor Analysis, Local Linear Embedding, Gene Clustering.
7. Recommendation systems, Tensor Decomposition, Autoencoder, WORD2VEC, DOC and COC.
8. Complex Networks – Erdos-Reynei, Scale-Free, Structure and dynamics, Laplacian Eigenmaps, handling sparse high dimensional data.
9. Time-series data, FFT, Time-Frequency Analysis, noise reduction, pulse frequency and phase detection, Poincare Plot, Cross-Correlation.

### ג. דרישות קדם:

יידע מוקדם: סטטיסטיקה והסתברות, מתמטיקה (חדו"א, מד"ר), יכולת תכנות בסיסית בשפת תכנות כלשהיא ושימוש בספריות שלה. אין דרישה לשפת תכנות ספציפית. הקורס יתמקד ב Matlab כשפת תכנות.

### ד. חובות / דרישות / מטלות:

בקורס זה יש חובת הגשה של כל תרגילי הבית.

ה. מרכיבי הציון הסופי:

הציון ייקבע לפי תרגילי הבית.

ו. ביבליוגרפיה:

ז. שם הקורס באנגלית:

Data science methods for physicists