**פרק 5 : מודלים מבוססי עצים**

חלק תיאורטי:

1. PCA

Read pages 559-570 in Bishop book – first part on PCA, and more in the internet ofcourse :)

1. Kernels:

Read about kernels in Bishop Chapter 6 up to 6.1 including

Solve exercise 6.3 in the end of the chapter

Now:

Read about KPCA in Bishop 586-590

Some intuition about kernels:

<https://sebastianraschka.com/Articles/2014_kernel_pca.html>

<https://arxiv.org/pdf/1207.3538.pdf>

1. Random Projection:

יש לקרוא את הפוסטבלוג הבא:

<https://machinelearningmedium.com/2017/07/28/random-projection-in-dimensionality-reduction/>

אשר סוקר מאמר בנושא הורדת מימדים ומשווה בין הביצועים של RP, PCA ומספר שיטות נוספות אשר נקרא עליהן בסעיף הבא.

ניתן להרחיב להסתכל על המאמר המקורי המדגיש את הבדלי החישובים הנדרשים בפועל לחישוב על אחת מהשיטות (מצורף כRP1).

ניתן להרחיב בנושא עם המאמר המקורי שמציג את RP(מצורף כRP2).

1. More Non Linear Methods:

* Read about the next non linear methods from the following references:
* Introduction to dimension reduction: LLE and ISO-MAP <https://people.eecs.berkeley.edu/~efros/courses/AP06/presentations/ThompsonDimensionalityReduction.pdf>
* Fast passage over the main algorithms

<https://www.math.uwaterloo.ca/~aghodsib/courses/f06stat890/readings/tutorial_stat890.pdf>

* Focus on ISOMAP with specific details

<http://www.robots.ox.ac.uk/~az/lectures/ml/tenenbaum-isomap-Science2000.pdf>

* Focus on laplacian Eigenmaps

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.9.5888&rep=rep1&type=pdf>

חלק מעשי:

בתרגיל זה נצלול לתרגיל קאגל של אתגר עם המון פיצ'רים.

הקאגל הבא: <https://www.kaggle.com/c/MerckActivity/overview>

מתאר את רמת הפעילות הכימית של מולקולות שונות בהינתן כ10000 פיצ'רים שונים המתארים אותם.

עליך לבנות רגרסור בעל דיוק מירבי לחיזוי הנ"ל.

בפתרונך עליך לכלול השלבים הבאים:

* אקספלורציה של המידע.
* ייצור פיצ'רים חדשים היכולים לעזור בעבודת המודל.
* בחינת מודלים
* השוואה בין המודלים שנבחנו.
* מסקנות.