

## תרגיל בית 2 במבוא ללמידה ממוכנת

מגשים:

ליאון גורין --- תז 214511214  
איתי בר --- תז 325839710

### תרגיל 1

1. נרצה להראות שקלסיפיקציה באמצעות חלונות parzen שווה

$$P(w_j|x) \geq P(w_i|x) \quad \forall i \neq j$$

נשתמש בכלל bayes:

$$P(x|w_j)P(w_j) \geq P(x|w_i)P(w_i) \quad \forall i \neq j$$

נשערך את ה likelihood  $P(x|w_j)$  באמצעות חלונות parzen:

$$P(x|w_j) = \frac{1}{n_j \cdot h^d} \sum_{k=1}^{n_j} \Phi\left(\frac{x_k^j - x}{h}\right)$$

בנוסף נחשב את ה prior  $P(w_j)$ , היחס בין מספר הדוגמאות בקטגוריה לכלל הדוגמאות:

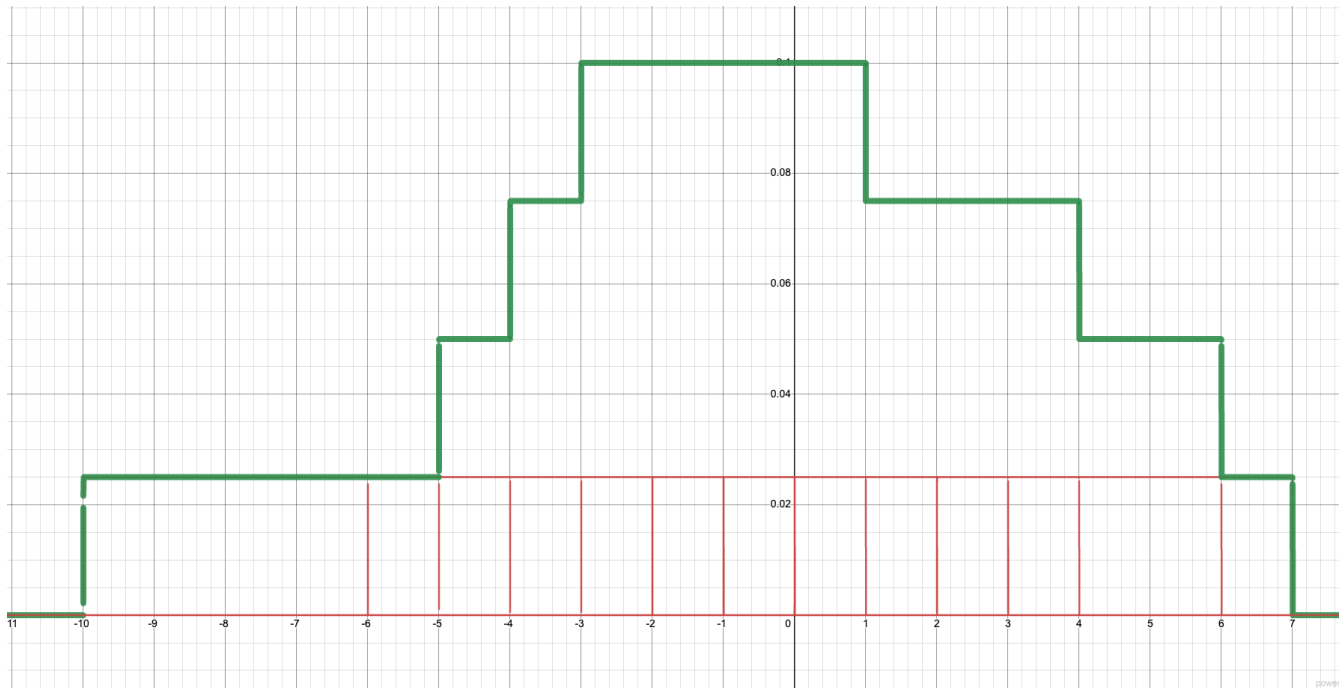
$$P(w_j) = \frac{n_j}{N}$$

נציב בחזרה ונקבל

$$\begin{aligned} \frac{1}{n_j \cdot h^d} \sum_{k=1}^{n_j} \Phi\left(\frac{x_k^j - x}{h}\right) \frac{n_j}{N} &\geq \frac{1}{n_i \cdot h^d} \sum_{k=1}^{n_i} \Phi\left(\frac{x_k^i - x}{h}\right) \frac{n_i}{N} \\ \sum_{k=1}^{n_j} \Phi\left(\frac{x_k^j - x}{h}\right) &\geq \sum_{k=1}^{n_i} \Phi\left(\frac{x_k^i - x}{h}\right) \end{aligned}$$

כנדרש

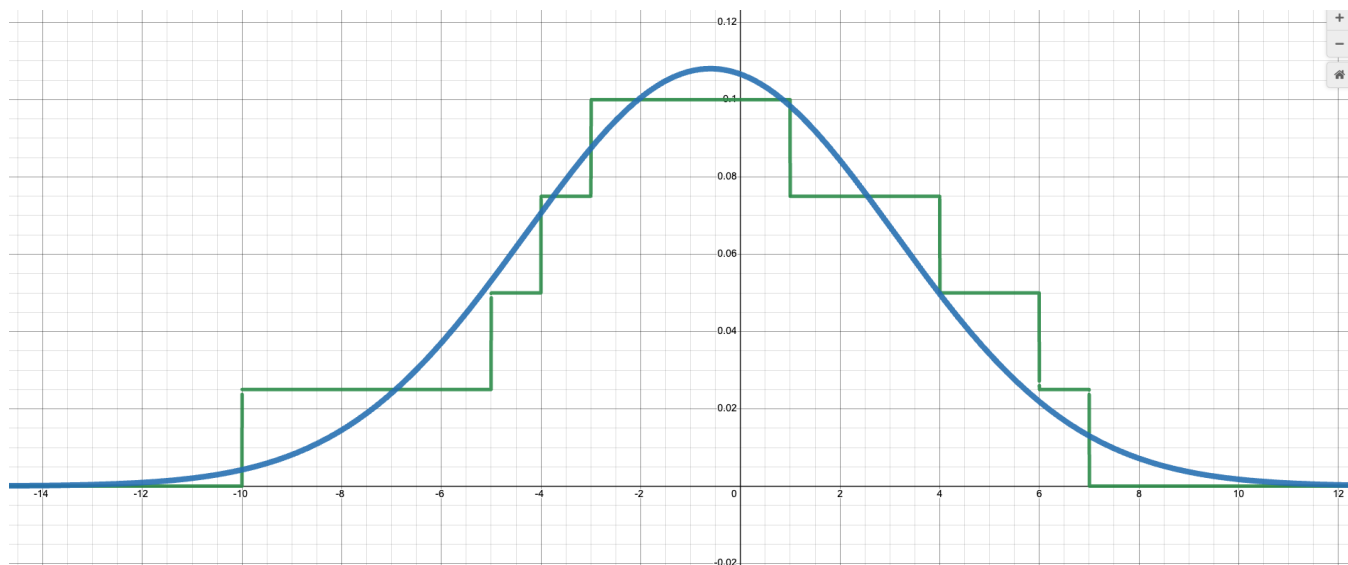
2.  
ה PDF שקיבלנו:



ההתפלגות נראית יחסית דומה להתפלגות נורמלית.  
נשערך באמצעות MLE את התוחלת וסטיית התקן:

$$\mu = \frac{1}{N} \sum_i^N x_k = \frac{1 + (-3) + 2 + 4 + 5 + (-8) + 0 + (-1) + (-2) + (-4)}{10} = -0.6$$

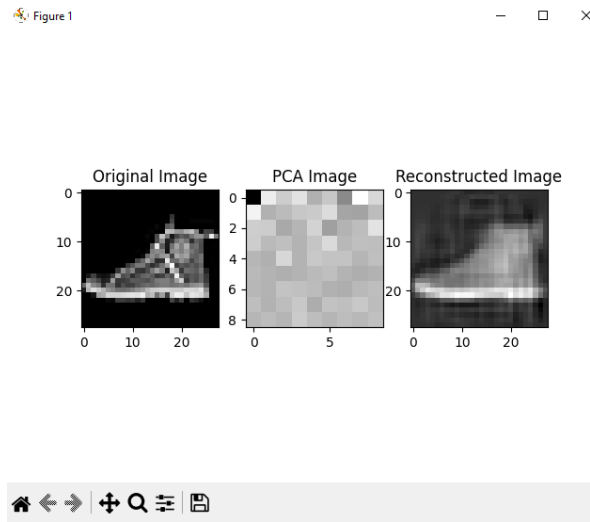
$$\sigma = \sqrt{\frac{1}{N} \sum_i^N (x_i - \mu)^2} = 3.69$$



ניתן לראות כי שני ה-PDF-ים אכן דומים.

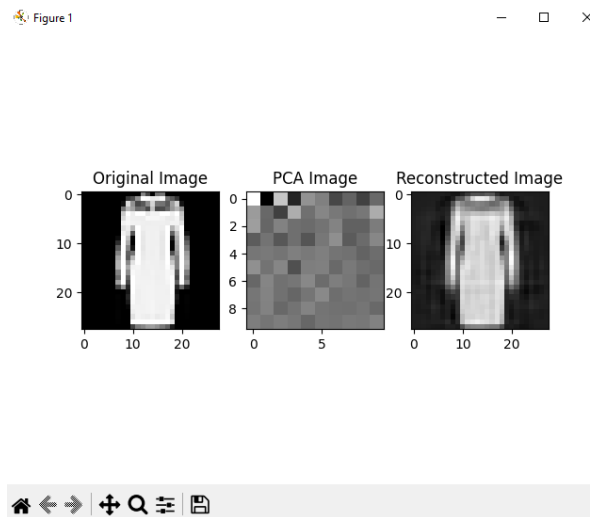
## תרגיל 2

3. הרצת PCA על ה train data עם מימד 81 (9 על 9):



4.

לפי הגרף אנחנו רוצים לחפש את מספר המימדים בעיכול גרף ה CDF כי לאחריו, אנחנו מקבלים תוצאות עם החזר מצתמצם ולכן כשהסתכלנו על הגרף הערכנו שהנקודה נמצאת בערך על מימד 100 (10 על 10):

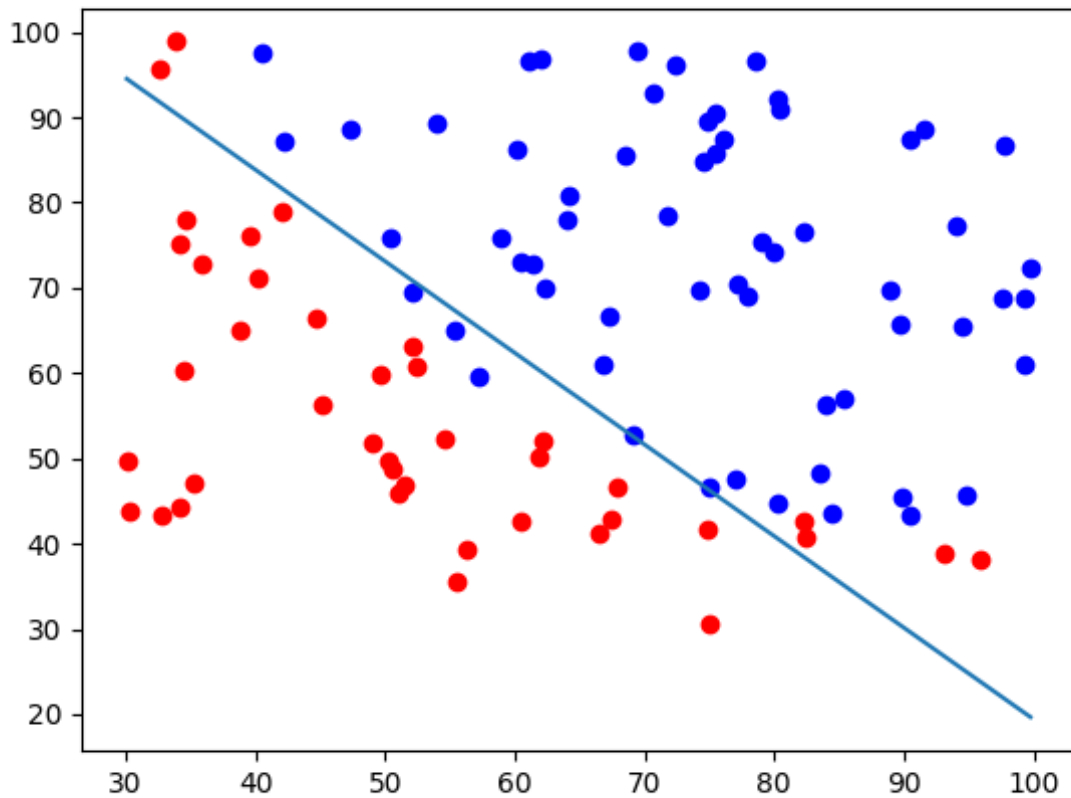


6. תוצאות שהשגנו על ידי הרצה עם מימדי PCA השווים ל-100, ועם k כפרמטר ל KNN ששווה ל-9:

```
PS D:\VISUAL STUDIO CODE\Machine-Learning-Homework\HW2\q2> python .\q2.py
Test accuracy is: 86.66%
```

### תרגיל 3

הדיוק הממוצע על סט ה test מתוך 10 הרצות שקיבלנו הוא 90%.  
להלן הקו המפריד הטוב ביותר על סט הטסט, על גבי כל הדאטא.



הדאטאסט שקיבלנו אינו linearly separable, מפני שלכל קו מפריד שנעביר יהיו outliers שלא יסווגו נכון.  
מסוג SVM לא לינארי יצליח להפריד את הדאטא.

Training data

