תרגיל בית שני- מבוא ללמידה ממוכנת

1. סעיף א':

<u>נראה שכל ה$classifation$ יכול להתפשט לחוק הבא:</u>

בהינתן $sample$ תשייך אליו את הקטגוריה אשר נותנת לו את רוב קולות של השכנים

כלומר בהינתן ה$sample$ j מתקיים שעבור כל $i \neq j$:

$$Pr\left(w_j \mid x \right) \geq Pr\left(w_i \mid x \right)$$

ונראה שמתקיים שלכל $i \in \{1, \dots, c\}$ מתקיים:

$$\sum_{k=1}^{n_j} \Phi\left(\frac{x_k^j - x}{h}\right) \geq \sum_{k=1}^{n_j} \Phi\left(\frac{x_k^i - x}{h}\right)$$

לפי חוק $bayes$ $Pr(A|B) = \frac{Pr(B|A) \cdot Pr(A)}{P(B)}$ מתקיים:

$$Pr\left(x \mid w_j \right) \cdot Pr\left(w_j\right) \geq Pr\left(x \mid w_i \right) \cdot Pr(w_i)$$

ולפי $parzen\ window$ אנחנו קובעים:

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

ובנוסף אנחנו יודעים שמתקיים:

$$Pr\left(w_j\right) = \frac{n_j}{n}$$

ועכשיו נפשט את האי שוויון שקיבלנו מ$bayes$:

$$Pr\left(x \mid w_j \right) \cdot Pr\left(w_j\right) \geq Pr\left(x \mid w_i \right) \cdot Pr(w_i)$$

$$\frac{1}{n_j} * \frac{1}{h^d} \sum_{k=1}^{n_j} \Phi\left(\frac{x_k^j - x}{h}\right) \cdot \frac{n_j}{n} \geq \frac{1}{n_i} * \frac{1}{h^d} \sum_{k=1}^{n_j} \Phi\left(\frac{x_k^i - x}{h}\right) \cdot \frac{n_i}{n}$$

$$\frac{1}{n} * \frac{1}{h^d} \sum_{k=1}^{n_j} \Phi\left(\frac{x_k^j - x}{h}\right) \geq \frac{1}{n} * \frac{1}{h^d} \sum_{k=1}^{n_j} \Phi\left(\frac{x_k^i - x}{h}\right)$$

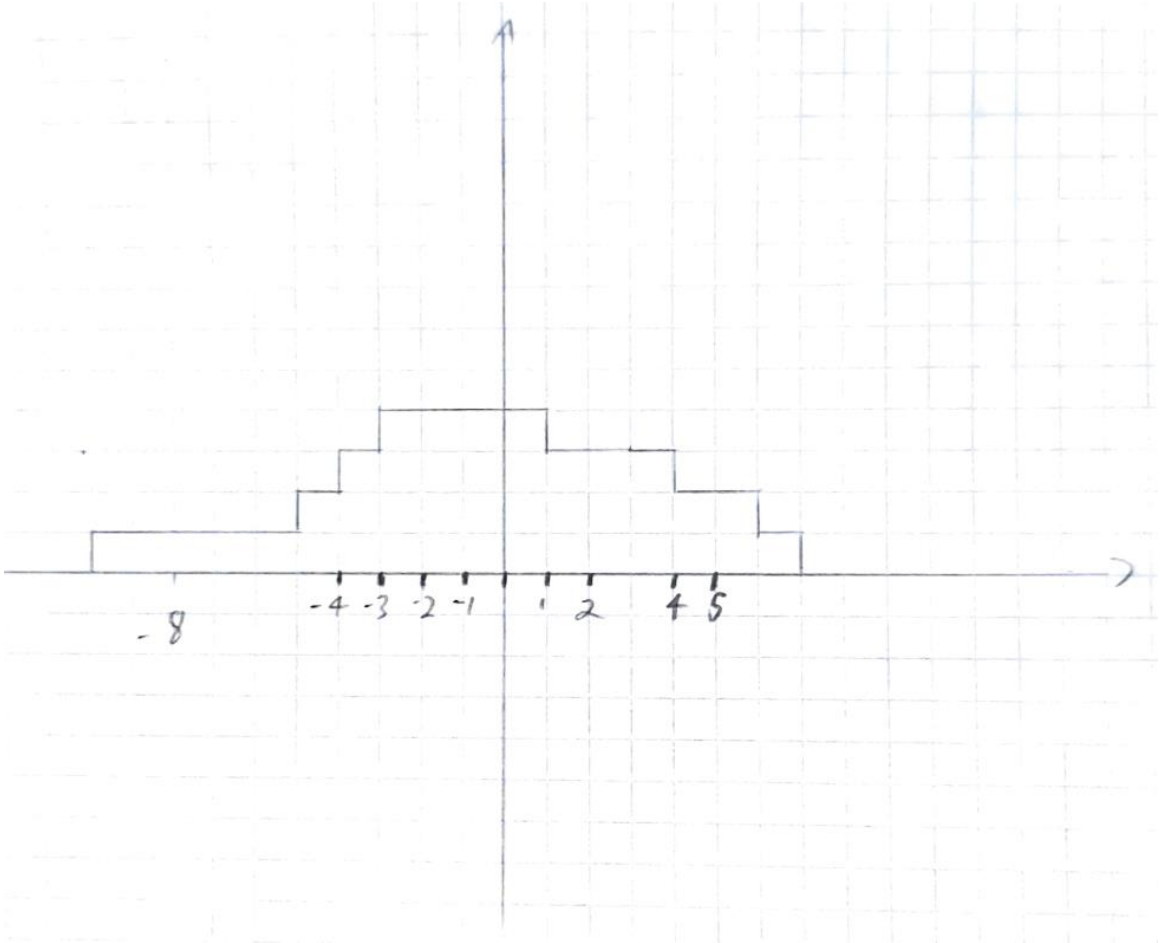
$$\sum_{k=1}^{n_j} \Phi\left(\frac{x_k^j - x}{h}\right) \geq \sum_{k=1}^{n_j} \Phi\left(\frac{x_k^i - x}{h}\right)$$

והוכחנו כנדרש את מה שהיינו צריכים להוכיח.

סעיף ב':

בהינתן ה$samples$ מהתפלגות שאינה ידועה $D = \{1, -3,2,4,5, -8,0, -1, -2, -4\}$, ובהינתן $h = 4$,

נשתמש פונקציית חלון בכדי לבנות PDF עבור ההתפלגות (בעזרת סרטוט):

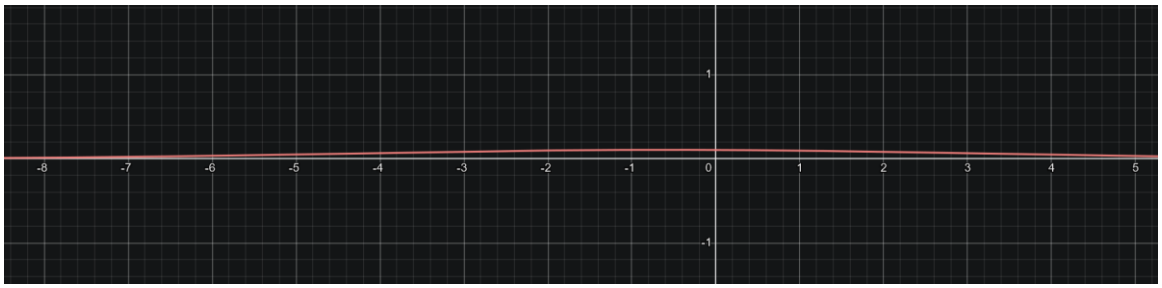


ובעזרת $desmos$ נבנה גאוסיין לדאטה, אבל לפני זה נחשב את התוחלת ואת ההתפלגות:

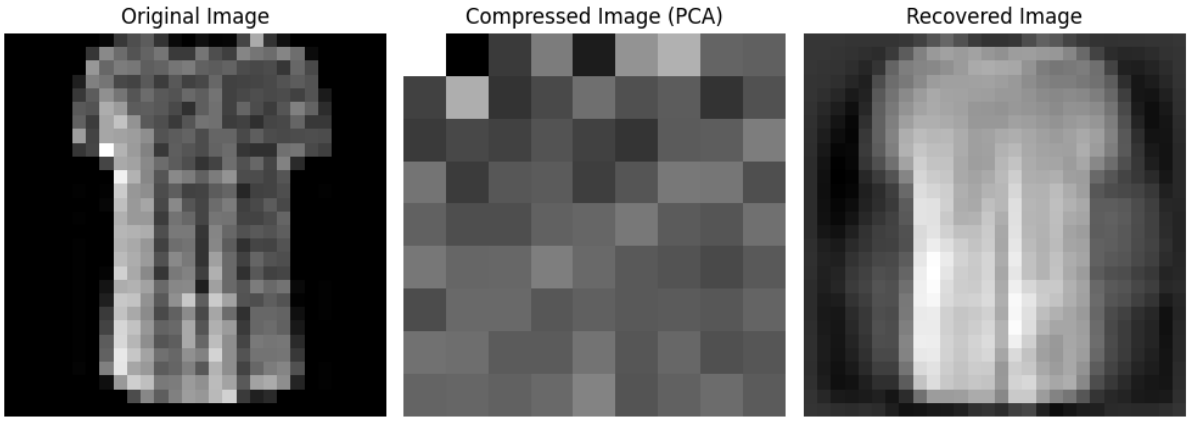
$$\mu = \frac{1 - 3 + 2 + 4 + 5 - 8 + 0 - 1 - 2 - 4}{10} = -0.6$$

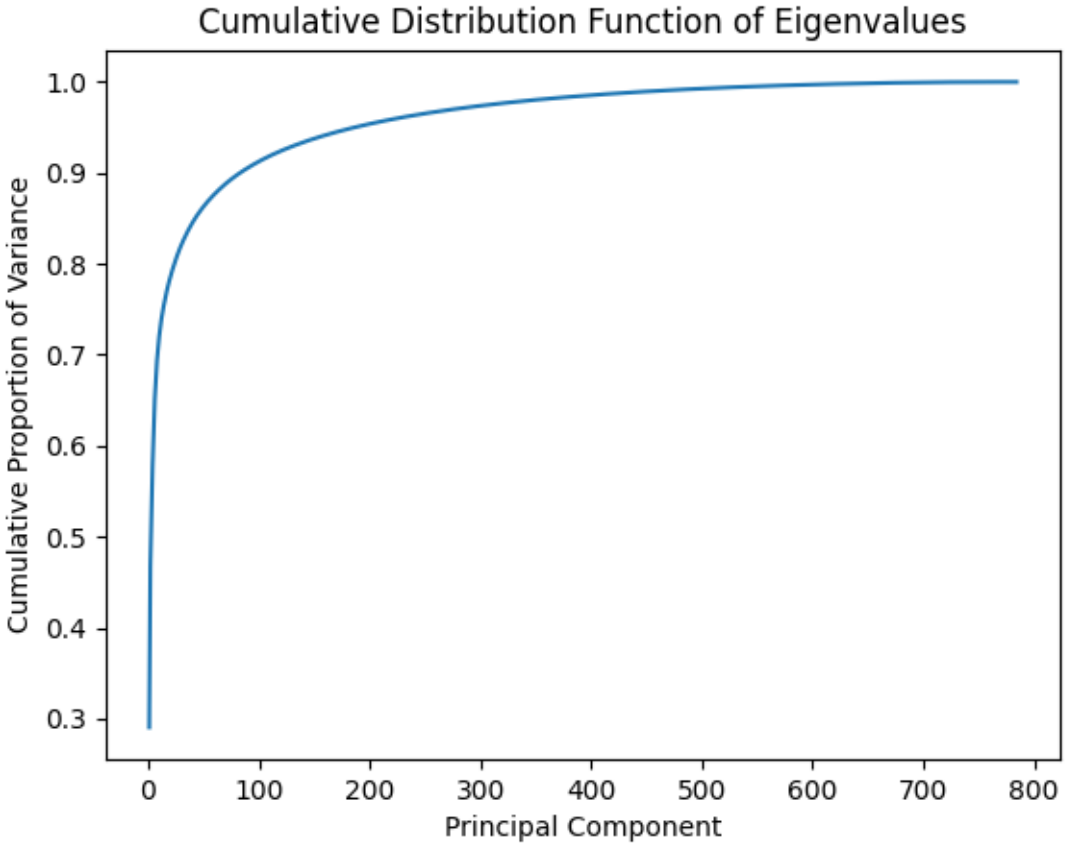
$$\sigma^2 = \frac{\sum_{i=1}^{10}(x_i - \mu)^2}{10} = 13.64$$

$$\sigma = 3.69$$



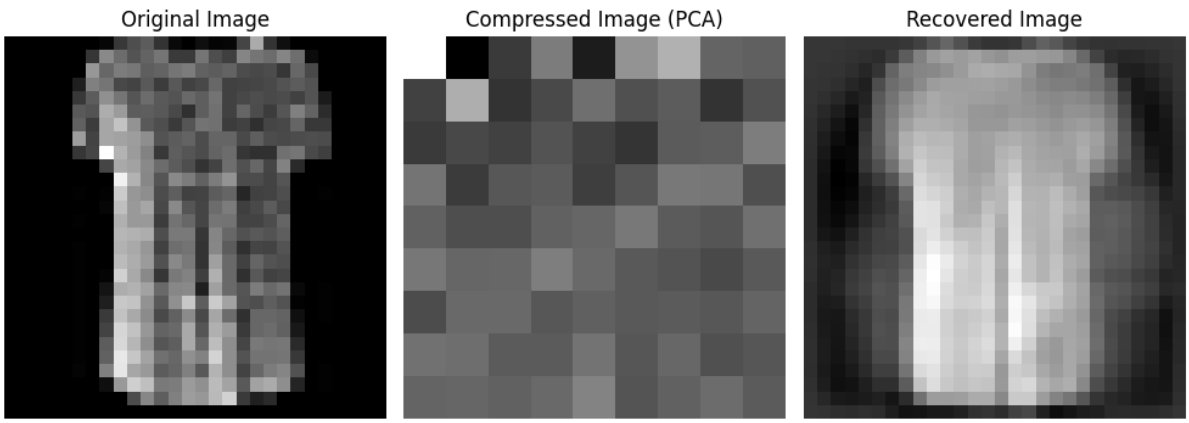
2. פתרון:





בחרנו להשאיר את המימד החדש ב81 מכיוון שעל פי הCDF שקיבלנו, החל מ$principal\ component = 81$ פונקציית הCDF מאטה את קצב העלייה שלה באופן משמעותי.

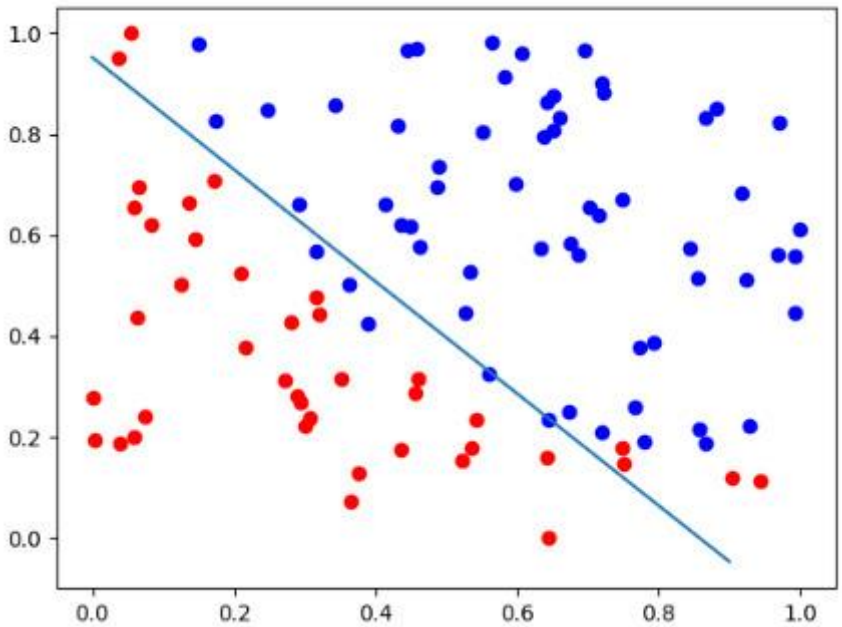
מכיוון שלא שינינו את גודל ה$principal\ components$ גם ה$plot$ מסעיף 3 נשאר אותו דבר:



לבסוף, קיבלנו עבור $k = 9$:

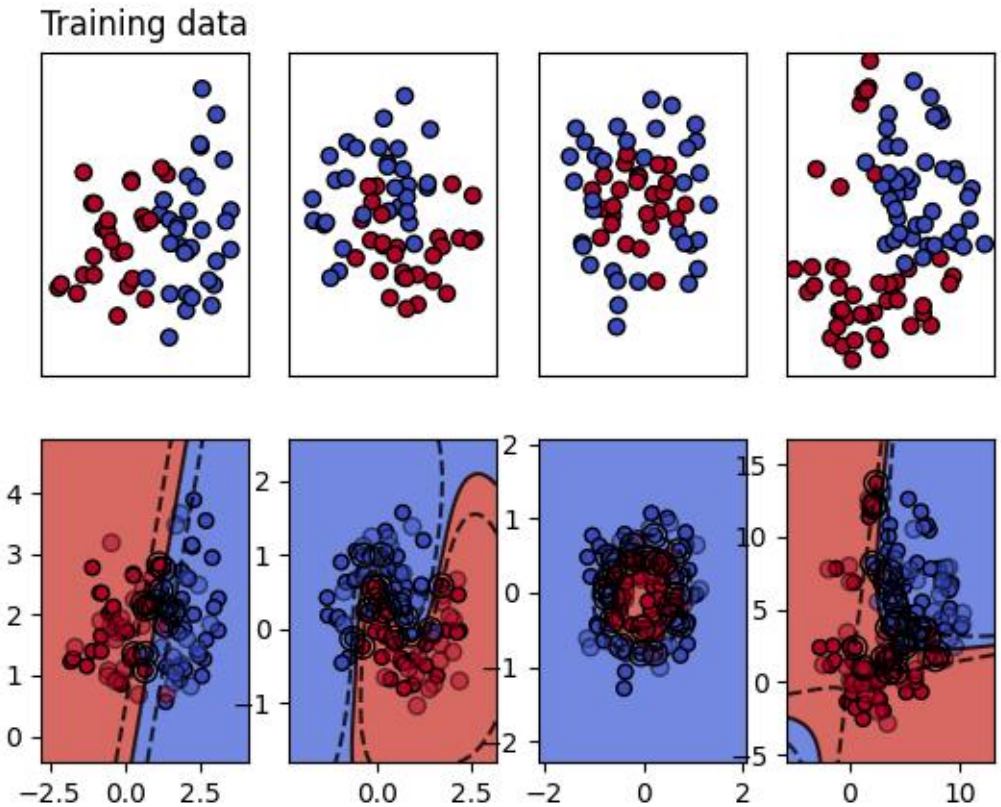
```
"C:\Users\2 ניר\Documents\GitHub\Machine_Learning_HW1\HW2\venv\Scripts\python.exe"
Test accuracy is: 86.6%

Process finished with exit code 0
```

3. פתרון:




```
Avg test accuracy: 90.69999999999997%
```

4. פתרון:




```
linear
0.95
rbf
0.975
rbf
0.95
poly
0.9655172413793104
```

א. בחרנו ב$linear\ kernel$.

בחרנו דווקא ב$kernel$ זה מכיוון שהתפלגות הנקודות נראית לינארית.

ב. בחרנו ב$rbf\ kernel$.

בחרנו דווקא ב$kernel$ זה מכיוון שניתן לראות ב$figure$ שהתפלגות הנקודות אינה פולינומית.

ג. בחרנו ב$rbf\ kernel$.

בחרנו דווקא ב$kernel$ זה מכיוון שניתן לראות ב$figure$ שהתפלגות הנקודות אינה פולינומית.

ד. בחרנו ב$poly\ kernel$.

בחרנו דווקא ב$kernel$ זה מכיוון שהתפלגות הנקודות לא נראית לינארית.