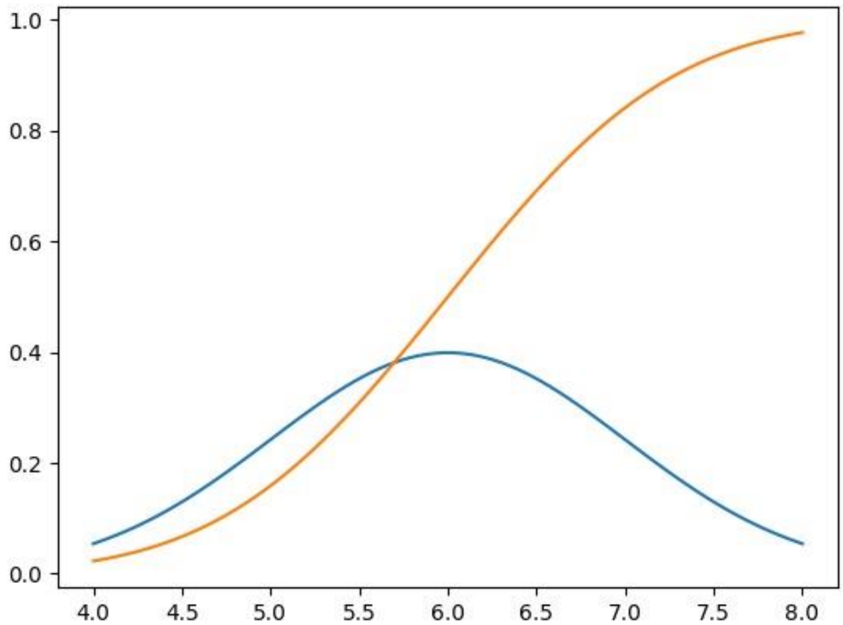
שאלה 5

5(a). בpdf המצורף.

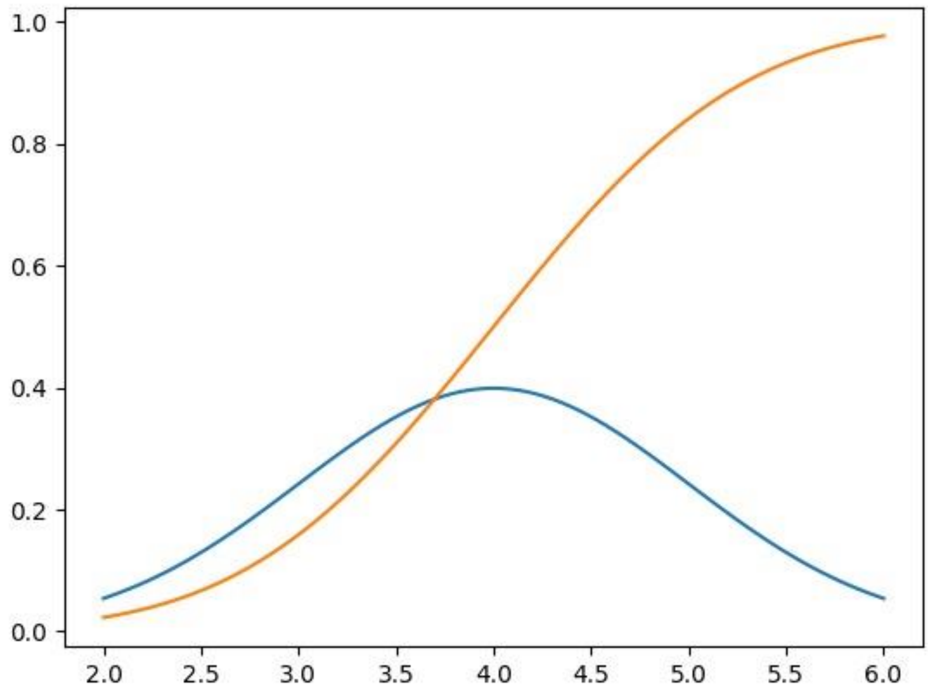
5(b).

1) הגרף הכתום מייצג את CDF , הגרף הכחול מייצג את PDF:

mean = 6



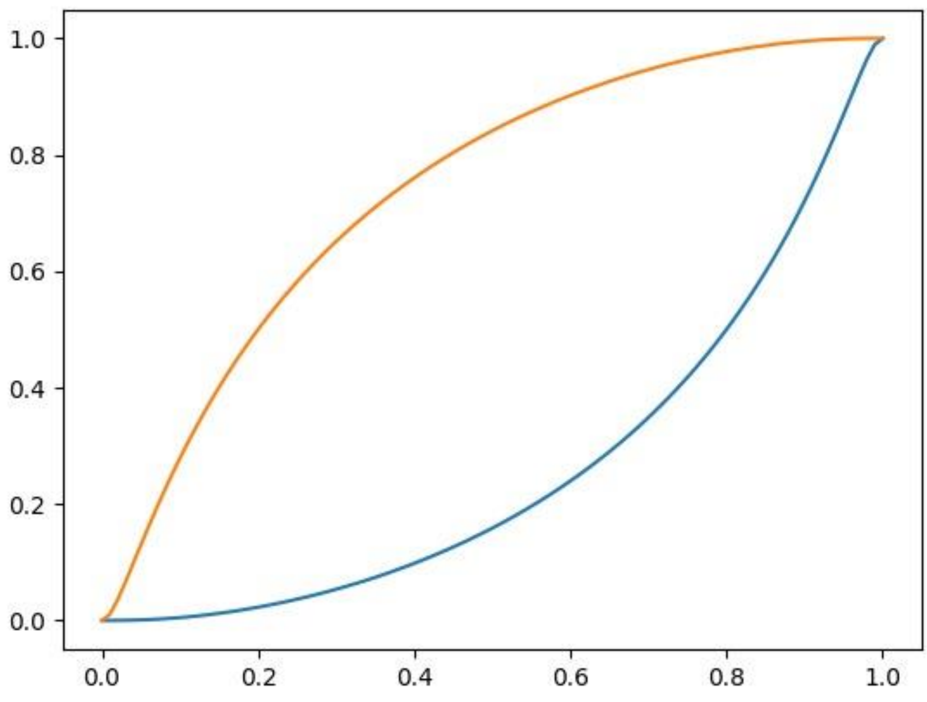
mean = 4



2) גרף CDF של h(x) כתלות ב x :

הגרף הכתום מיצג כאשר mean = 4

הגרף הכחול מיצג
כאשר mean = 6

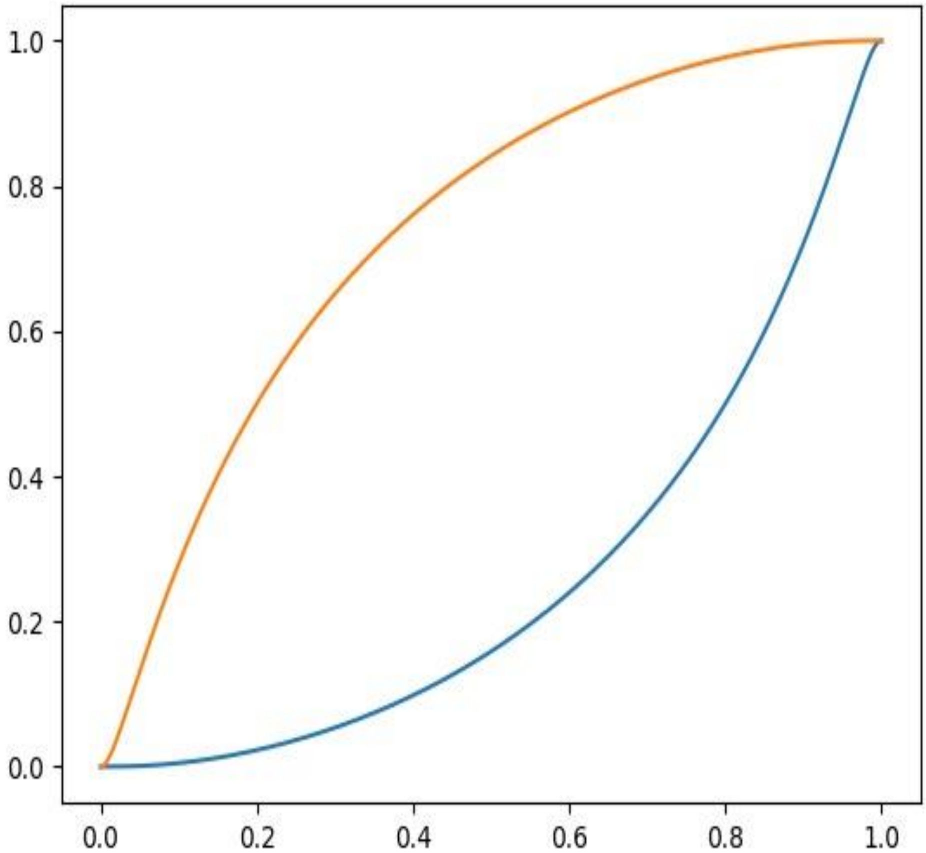


(3

גרף CDF המצטברת של h(x) כתלות ב x :

הגרף הכתום מיצג כאשר mean = 4

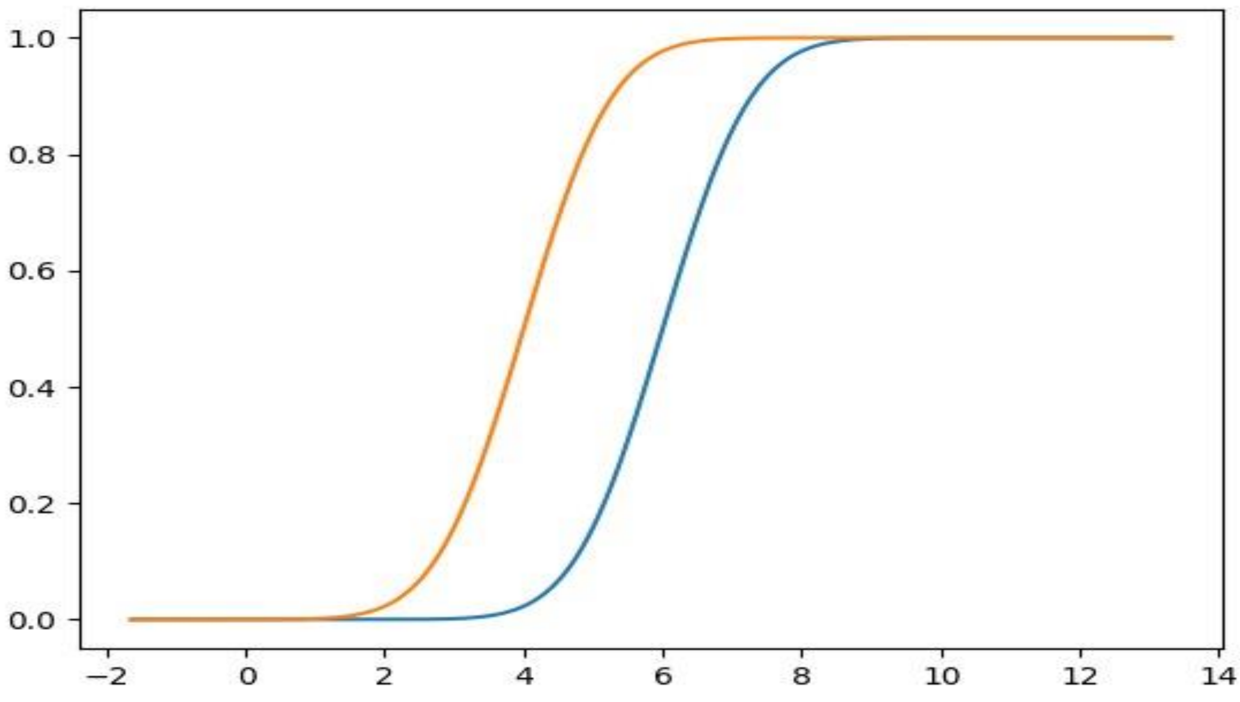
הגרף הכחול מיצג כאשר mean = 6



(4

הגרף הכתום מיצג כאשר mean = 4

הגרף הכחול מיצג כאשר mean = 6



(5

fpr at 0.2 is: 0.5

tpr at 0.2 is: 0.9772498680518208

fpr at 0.4 is: 0.2396220825571478

tpr at 0.4 is: 0.9019047325061418

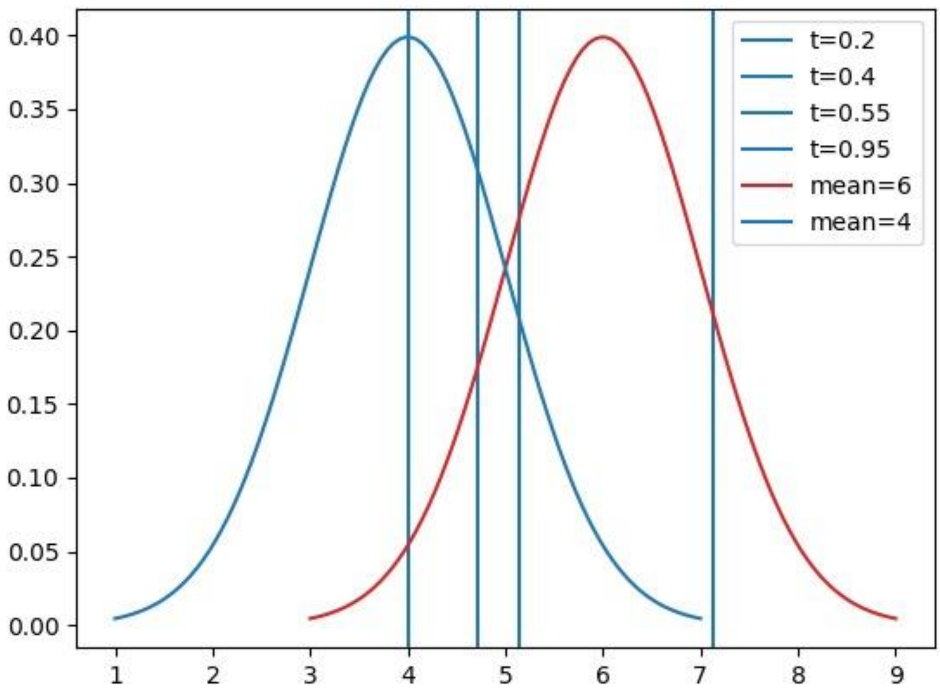
fpr at 0.55 is: 0.12615568092951324

tpr at 0.55 is: 0.8037927037201944

fpr at 0.95 is: 0.000892162169381483

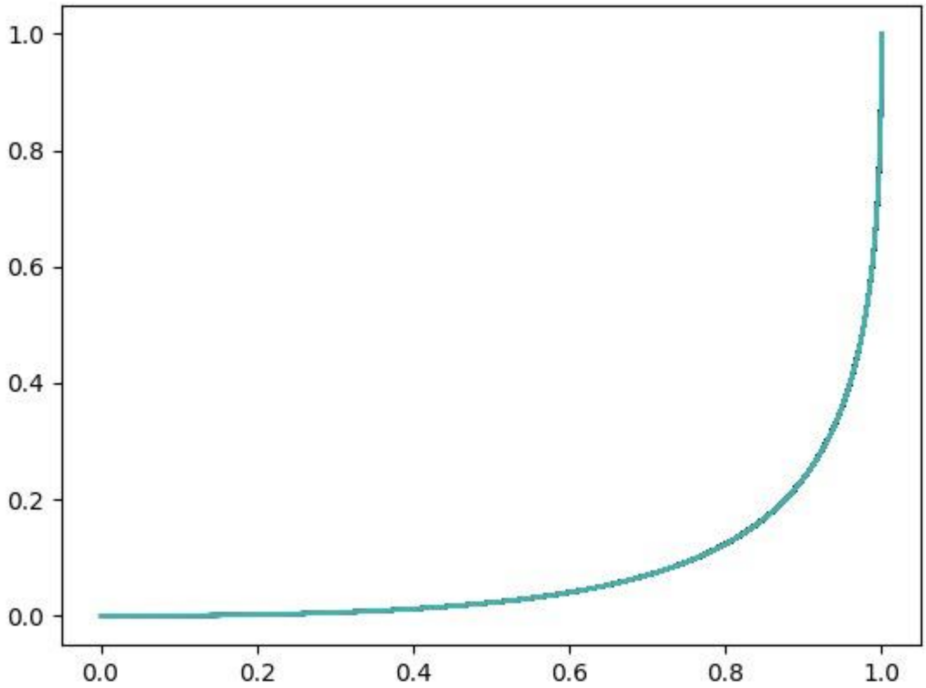
tpr at 0.95 is: 0.13051420087337862

(6



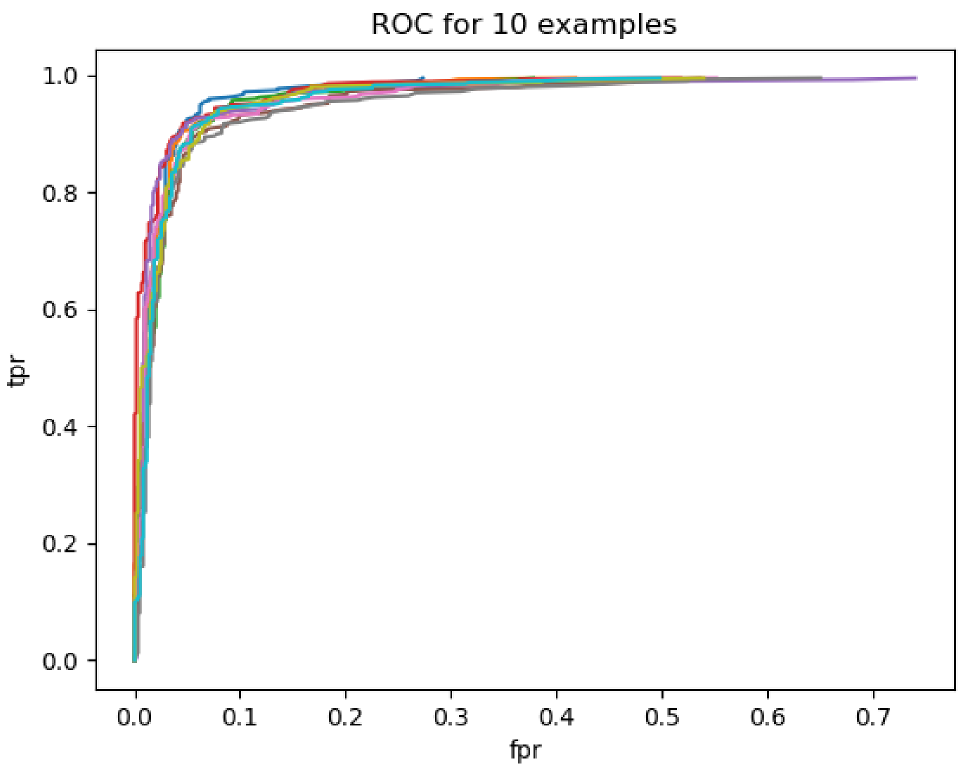
t=0.2
t=0.4
t=0.55
t=0.95
mean=6
mean=4
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1
2
3
4
5
6
7
8
9

7) בציר X מופיעים ערכי TPR ובציר ה y מיוצגים FPR -



שאלה 7

ב- עבור כל איטרציה צבע שונה -



בס״ד

313308160 לירוי קלירמן

תרגיל 3 - IML

LinearRegression...

1. a) $\sigma > 0 \in \mathbb{R}$.

$y \mid x = x$ מתפלג $N(h(x), \sigma^2)$ כיוון ש-m הדגימות בת"ל מתקיים:

$$L_S(h) = \prod_{i=1}^{m} l_{\text{likelihood}}(h, (x_i, y_i)) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y_i - h(x_i))^2}{2\sigma^2}\right)$$

b) עלינו להראות כי

$$\underset{h\in H}{\operatorname{argmin}} L_S(h) = \underset{h\in H}{\operatorname{argmax}} L_S(h)$$

$$\left(L_S(h) = \frac{1}{m}\sum_{i=1}^{m}(h(x_i) - y_i)^2\right)$$

$$L_S(h) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - h(x_i))^2}{2\sigma^2}\right) =$$

$$m \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\sum_{i=1}^{m}(y_i - h(x_i))^2}{2\sigma^2 m}\right) \quad \#$$

נשים לב כי מה שמשתנה בעקבות h זה $\exp\left(-\sum_{i=1}^{m}(y_i - h(x_i))^2\right)$, כלומר $\sum_{i=1}^{m}(y_i - h(x_i))^2$ רק $\Leftrightarrow \exp\left(-\sum_{i=1}^{m}(y_i - h(x_i))^2\right)$

מכיון שכל הביטוי # גדל. מכאן, שמתקיים:

$$\underset{h\in H}{\operatorname{argmax}} L_S(h) =$$

$$\underset{h\in H}{\operatorname{argmax}} = \underset{h\in H}{\operatorname{argmax}} \frac{m}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\sum_{i=1}^{m}(y_i - h(x_i))^2}{2\sigma^2 m}\right)$$

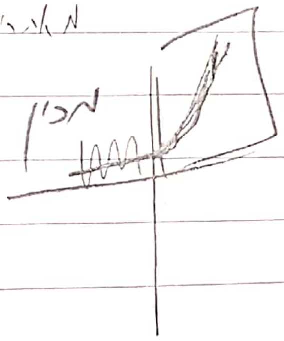
$\Leftrightarrow$ מהמעבר הקודם

$$\underset{h\in H}{\operatorname{argmax}} \exp\left(-\sum_{i=1}^{m}(y_i - h(x_i))^2\right)$$

$$\Updownarrow$$

$$\underset{h\in H}{\operatorname{argmax}} \frac{1}{e^{\sum_{i=1}^{m}(y_i - h(x_i))^2}}$$

מכיון שהביטוי חיובי עלינו להקטין כדי להגדיל סה״כ את הביטוי



$$\Updownarrow$$

$$\underset{h\in H}{\operatorname{argmin}} \sum_{i=1}^{m}(y_i - h(x_i))^2$$

$$\Updownarrow$$

$$\boxed{\underset{h\in H}{\operatorname{argmin}} L_S(h)}$$

) מקבל!

∎

ROC Curve and Random classification

3) a) לכל t חוקי:

$$p = TPR(t) = P_D(h_t(x) = 1 \mid y = 1)$$

$$= P_D(z \geq t \mid y = 1) = \frac{P_D(y = 1 \wedge z \geq t)}{P_D(y = 1)}$$

z מתפלג אחיד כמו x

אז:

$$\frac{\cancel{P_D(y=1)} \cdot P_D(z \geq t)}{\cancel{P_D(y=1)}}$$

אז:

$$p = \int_{x=t}^{1} 1 \, dx$$

$$= x \Big|_{t=x}^{1} = \boxed{1 - t}$$

$$\boxed{t = 1 - p}$$

b)

$$FPR(t) = P_D(h_t(x) = 1 \mid y = 0)$$

מהסעיף הקודם כי $TPR(t) = p$ אז $t = 1 - p$:

$$FPR(t) = P_D(h_t(x) = 1 \mid y = 0) = P_D(h_t(x) = 1 \mid y = 0) =$$

$$\frac{P_D(z < t, y = 0)}{P_D(y = 0)} = \frac{\cancel{P_D(z < t) \cdot P_D(y = 0)}}{\cancel{P_D(y = 0)}} = P_D(z < t)$$

$$= \int_{0}^{t=x} 1 \, dx = x \Big|_{0}^{t} = \underline{\underline{t}}$$

וגם $t = 1 - p$

נקבל: $FPR(1 - p) = \boxed{1 - p}$

202

לא להגיש

4) (a) יהי $x \in [0,1]$ מתפלג אחיד

$$y(x)=\begin{cases}1 & x\geq 0.5\\ 0 & \text{else}\end{cases}$$

ומסווגים $h(x)=1$ $x>0.6$

$h(x)=0$ $x<0.5$

$h(x)=x$ $0.5<x<0.6$ אזי אם $0.5<t\leq 0.6$ נקבל

$$\begin{cases} h_t(x)=1/0 & 0.5<x<0.6 \\ h_t(x)=1 & 0.6<x<1 \\ h_t(x)=0 & 0<x<0.5 \end{cases}$$

ומכיוון שמתקיים כי FPR לא משתנה מאחר ועבור $y=0$ מתקיים כי $x\leq 0.5 \Rightarrow h_t(x)=0 \Rightarrow FPR=0$

נשים לב כי TPR עולה כאשר $t \to 0.5$ מכיוון שלכל מספר הולך וקטן מקיים $h_t(x)=1$ עבור כאשר $0.5<x<0.6$ נסמן $h(x)=1$ $0.5<x<0.6$ ולכן ה-TPR עולה.

– כעת נראה דוגמא בה FPR משתנה ו-TPR קבוע.

נקח כמו אותו דבר למעט

$$y(x)=\begin{cases}1 & x\geq 0.6\\ 0 & \text{else}\end{cases}$$

← נראה לשמור את הדוגמא t:

אם $y(x)=1 \Rightarrow 1\geq x\geq 0.6$ וגם $h(x)=1$ ולכן $h_t(x)=1 \Rightarrow$ TPR קבוע

אם $y(x)=0$ נסמן $x<0.6$, אם $x\leq 0.5$ אזי $h(x)=0 \Rightarrow h_t(x)=0$

ואם $0.5<x<0.6$ $h(x)=x \Rightarrow$ כאשר t עולה מס' הדוגמאות שמקיימים $h_t(x)=0$ [illegible]

ולכן מס' הדוגמאות המקיימים $h_t(x)=1$ עולה ולכן FPR עולה כנדרש.

(b) יהי t, נסמן $t+\varepsilon$ עבור $\varepsilon>0$.

עבור כי
$$h_t(x)=\begin{cases}1 & h(x)\geq t\\ 0 & h(x)<t\end{cases}$$

עבור $t+\varepsilon$:

$$\left|\{x \mid h_t(x)=1\}\right| \geq \left|\{x \mid h_{t+\varepsilon}(x)=1\}\right|$$

$$FPR=\frac{1}{N_0}\sum_{i:y=0} 1_{h_t(x)=1} \qquad TPR=\frac{1}{N_1}\sum_{y_i=1} 1_{h_t(x_i)=1}$$

נציג מחדש את ה-h ו-t:

$y_i=0/1$ אזי. ולכן $\sum 1_{h_{t+\varepsilon}(x_i)=1} \leq \sum 1_{h_t(x_i)=1}$ – ולכן

$$FPR(t)=\frac{1}{N_0}\sum_{y_i=0} 1_{h_t(x_i)=1} \geq \frac{1}{N_0}\sum_{y_i=0} 1_{h(t+\varepsilon)(x_i)=1}$$ ומכאן

$$\Downarrow$$

$$FPR(t+\varepsilon)\leq FPR(t)$$

$$\frac{1}{N_0}\sum_{y_i=1} 1_{h_t(x_i)=1} \geq \frac{1}{N_0}\sum_{y_i=1} 1_{h_{t+\varepsilon}(x_i)=1}$$ ומכאן

$$\Downarrow$$

$$TPR(t+\varepsilon)\leq TPR(t)$$

כלומר קיבלנו כי ככל ש-t גדל, יש לנו פחות נקודות. ומקבלים פונקציה מונוטונית עולה.

5 a) LDA $\Longrightarrow$ $D(X|y=y)=\mathcal{N}(\mu_y,\Sigma)$ ולכן מתקיים:

$$h_D(x)=\operatorname{argmax}_y\ x^T\Sigma^{-1}\mu_y-\frac{1}{2}\mu_y^T\Sigma^{-1}\mu_y+\log(\pi_y)$$ וכן $y\in\{0,1\}$

$$h_D(x)=\operatorname{sign}\left(x^T\Sigma^{-1}\mu_1-\frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1+\log\pi_1-\left(x^T\Sigma^{-1}\mu_0-\frac{1}{2}\mu_0^T\Sigma^{-1}\mu_0+\log\pi_0\right)\right)$$

(לקחת את הסימן של ההפרש של הערך של שני הביטויים. כאשר הביטוי של $y=1$ גדול יותר מקבלים חיובי ואז h מחזיר $y=1$ ואם לא $y=0$, ולכן הנוסחה נכונה)

$$\Longrightarrow \operatorname{sign}\left(x^T\underbrace{\left(\Sigma^{-1}\mu_1-\Sigma^{-1}\mu_0\right)}_{\#}-\underbrace{\frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1+\log\pi_1-\frac{1}{2}\mu_0^T\Sigma^{-1}\mu_0-\log\pi_0}_{*}\right)$$

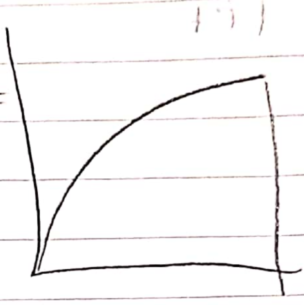
אם נסמן כל ביטוי ב-$*$ ו-#

נשים לב כי $*$ זה מס' חופשי ו-# זה המקדם של x.

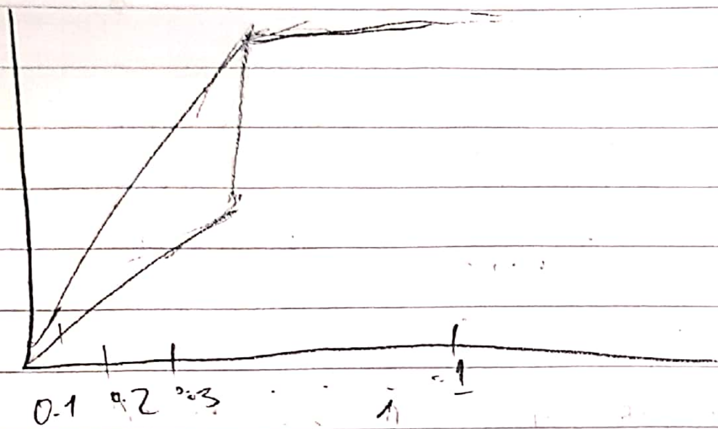
ולכן נגדיר $x'=\begin{pmatrix}x\\1\end{pmatrix}$, $w=\begin{pmatrix}\#\\ *\end{pmatrix}$ כמו שהגדרנו: $\operatorname{sign}(x\cdot\# - *)$

ולכן $h_D=\operatorname{sign}(x\cdot x, w)$ מקיימת הנ"ל!

6) a)

$h_1 =$ 

, לכן קווית הוכחה:



נקח ROC_1 כך ש- $TPR = 2t$ $0 < t < \frac{1}{2}$

$TPR = 1$ $\frac{1}{2} < t < 1$

$FPR = t$ לכל t

לכן ש $h_1 = AUC_1 = \boxed{\frac{3}{4}}$

נקח ROC_2 כך ש- $TPR = t$ $0 < t < \frac{11}{24}$

$TPR = 1$ $\frac{11}{24} < t < 1$

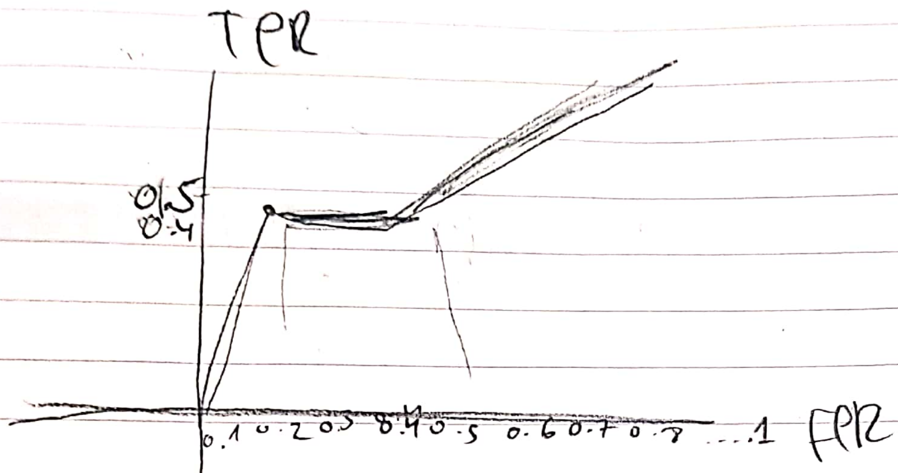
$FPR = t$ לכל t

$$h_2 = AUC_2 = \frac{121}{1152} + \frac{13}{24} = \boxed{0.64}$$

אמנם, משקיפים $\underline{h_1 > h_2}$ אך עבור $\frac{11}{24} < t < 0.5$ ROC_2 יותר טוב מ-ROC_1 כיוון שמעליה כמו ה-TPR גבוה יותר וכן FPR נמוך

b) לכן ROC_2 יותר טוב מ-ROC_1 כי נקודה זו עבור t, FPR, ה-TPR ב-ROC_2 יותר מ-ROC_1 וכן עבור t זה FPR נמוך יותר TPR
ROC_2 יותר מ- ROC_2, לכן נעדיף את $\underline{ROC_1!}$

6ג



בקטע $[0.2,0.5]$ TPR קבוע כי FPR גדל ולכן $t\in 0.5$ 0.2
כי יגדיל את TPR ורק יגדיל את FPR ולכן לא נבחר t בקטע הזה
עבור $t\notin[0.2,0.5]$ גם TPR גם FPR יגדלו ולכן יהיה שינוי בבחירת t.

6א) ב-LDA וב-logistic regression אנו מניחים את ההתפלגות y|x ולא x|y
ובגלל שמודלים את ... ולכן יש לו סיכוי של ... את D ולא מניחים כגרסה
אחרי רק את ה-x, (ההתפלגויות המודל) לכן אנחנו ב-G את ההסתפקות D ומניחים דיוק
ב-generative ... לא היה צריך ב-D את ההתפלגות הזו.