מבוא לחישוביות עצבית – עבודה מסכמת

ניר כפרי – 313160194

אביב אזמנוז – 204805469

חלק 1.

א.

$$\epsilon = \frac{1}{2}(r - w_1u_1 - w_2u_2)^2$$
$$-\eta\frac{\partial\epsilon}{\partial w_1} = \eta(r - w_1u_1 - w_2u_2) * u_1 = \eta(w_1u_1^2 + w_2u_1u_2 - ru_1)$$
$$-\eta\frac{\partial\epsilon}{\partial w_2} = \eta(r - w_1u_1 - w_2u_2) * u_2 = \eta(w_2u_2^2 + w_1u_1u_2 - ru_2)$$
$$\mathrm{w}_1^{\mathrm{i}} = \mathrm{w}_1^{\mathrm{i}-1} + \eta(r - w_1u_1 - w_2u_2) * u_1 = \mathrm{w}_1^{\mathrm{i}-1} + \eta(ru_1 - w_1u_1^2 - w_2u_1u_2)$$
$$= \mathrm{w}_1^{\mathrm{i}-1} + \eta(\alpha_1 - w_1\gamma_1 - w_2\beta)$$
$$\mathrm{w}_2^{\mathrm{i}} = \mathrm{w}_2^{\mathrm{i}-1} + \eta(r - w_1u_1 - w_2u_2) * u_2 = \mathrm{w}_2^{\mathrm{i}-1} + \eta(-w_2u_2^2 - w_1u_1u_2 + ru_2)$$
$$= \mathrm{w}_2^{\mathrm{i}-1} + \eta(\alpha_2 - w_2\gamma_2 - w_1\beta)$$

B.1.

$$y = w_1u_1 + w_2u_2$$
$$\langle ru_1\rangle = \alpha_1 = \alpha, \langle ru_2\rangle = \alpha_2 = 2\alpha$$
$$\langle u_1^2\rangle = \gamma_1 = \gamma\,, \langle u_2^2\rangle = \gamma_2 = 2\gamma$$
$$\langle u_1u_2\rangle = \beta = 0$$
$$-w_1u_1^2 - w_2u_1u_2 + ru_1 = -w_1\gamma - w_2\beta + \alpha \rightarrow\ \beta = 0 \rightarrow\ w_1\gamma = \alpha \rightarrow w_1 = \frac{\alpha}{\gamma}$$
$$-w_2u_2^2 - w_1u_1u_2 + ru_2 = 0 \rightarrow -2\mathrm{w}_2\gamma - \mathrm{w}_1\beta + 2\alpha = 0 \rightarrow\ \beta = 0 \rightarrow 2\mathrm{w}_2\gamma = 2\alpha \rightarrow \mathrm{w}_2\gamma = \alpha$$
$$\rightarrow \mathrm{w}_2 = \frac{\alpha}{\gamma}$$

$$y = \frac{\alpha}{\gamma}u_1 + \frac{\alpha}{\gamma}u_2 = \frac{\alpha}{\gamma}(\mathrm{u}_1 + \mathrm{u}_2)$$

2.

$$\langle ru_1\rangle = \alpha_1 = \alpha, \langle ru_2\rangle = \alpha_2 = \alpha$$
$$\langle u_1^2\rangle = \gamma_1 = \gamma\,, \langle u_2^2\rangle = \gamma_2 = \gamma$$
$$\langle u_1u_2\rangle = \beta > 0$$

$$\frac{\partial\epsilon}{\partial w_1} = -w_1u_1^2 - w_2u_1u_2 + ru_1 = 0 \rightarrow \alpha - w_2\beta - w_1\gamma = 0 \rightarrow w_1 = \frac{\alpha - w_2\beta}{\gamma}$$

$$\frac{\partial\epsilon}{\partial w_2} = -w_2u_2^2 - w_1u_1u_2 + ru_2 = 0 \rightarrow \alpha - w_1\beta - w_2\gamma = 0 \rightarrow \alpha - \frac{\alpha - w_2\beta}{\gamma}\beta - w_2\gamma = 0$$
$$\rightarrow \frac{\alpha\gamma - \alpha\beta + w_2\beta^2 - w_2\gamma^2}{\gamma} = 0 \rightarrow if\ \gamma \neq 0, w_2 = \frac{\alpha(\beta - \gamma)}{(b^2 - \gamma^2)} \rightarrow\ w_2 = \frac{\alpha(\beta - \gamma)}{(\beta - \gamma)(\beta + \gamma)}$$
$$\rightarrow w_2 = \frac{\alpha}{\beta + \gamma}$$

$$w_1 = \frac{\alpha - w_2\beta}{\gamma} \rightarrow\ w_1 = \frac{\alpha - \frac{\alpha}{\beta + \gamma}\beta}{\gamma} \rightarrow w_1 = \frac{\alpha\beta + \alpha\gamma - \alpha\beta}{\gamma(\beta + \gamma)} = \frac{\alpha}{\beta + \gamma}$$

$$y = w_1 u_1 + w_2 u_2 \rightarrow\ y = \frac{\alpha}{\beta+\gamma} u_1 + \frac{\alpha}{\beta+\gamma} u_2 \rightarrow y = (\mathrm{u}_1 + \mathrm{u}_2)\frac{\alpha}{\beta+\gamma}$$

C.1.

$$y = wz$$
$$z = j_1 u_1 + j_2 u_2$$
$$y = w(j_1 u_1 + j_2 u_2) = w j_1 u_1 + w j_2 u_2$$
$$\epsilon = \frac{1}{2}(r-y)^2 = \frac{1}{2}\left(r - w(j_1 u_1 + j_2 u_2)\right)^2 = \frac{1}{2}(r - w j_1 u_1 - w j_2 u_2)^2 = \frac{1}{2}(r - wz)^2$$
$$-\eta \frac{\partial \epsilon}{\partial w} = (r - w j_1 u_1 - w j_2 u_2)(j_1 u_1 + j_2 u_2) = r j_1 u_1 + r j_2 u_2 - w j_1^2 u_1^2 - 2 w j_1 j_2 u_1 u_2 - w j_2^2 u_2^2$$
$$-\eta\left(\frac{1}{2}(r - wz)^2\right)' = -\eta\frac{\partial \epsilon}{\partial w} = \eta z(r - wz) = \eta(wz^2 - rz) = 0 \rightarrow w = \frac{\eta r z}{\eta z^2} \rightarrow\ if\ z > 0\ \&\ \eta \neq 0$$
$$\rightarrow \Delta w = \frac{r}{z}.$$
$$w_i = w_{i-1} + \frac{rz}{z^2}$$

2.

נכניס את המשתנים למטריצה (במקרה של PCA מטריצת הקוווריאנס):

$$C = \begin{pmatrix} \gamma_1 & \beta \\ \beta & \gamma_2 \end{pmatrix}$$

בעבור מקרה 1:

$$(r u_1) = \alpha_1 = \alpha, (r u_2) = \alpha_2 = 2\alpha$$
$$(u_1^2) = \gamma_1 = \gamma\ , (u_2^2) = \gamma_2 = 2\gamma$$
$$(u_1 u_2) = \beta = 0$$

$$C = \begin{pmatrix} \gamma & 0 \\ 0 & 2\gamma \end{pmatrix}$$

Linear network:

$$y = \frac{\alpha}{\gamma}(\mathrm{u}_1 + \mathrm{u}_2)$$

PCA:

$$y = wz$$

$$w = \eta \frac{rz}{z^2}$$

$$\frac{rz}{z^2} = \frac{r(j_1 u_1 + j_2 u_2)}{(j_1 u_1 + j_2 u_2)^2} \rightarrow j_1 = 0, j_2 = 1 \rightarrow \frac{r u_2}{u_2^2} = \frac{\alpha}{\gamma}$$

$$C.\, y = wz = \frac{\alpha}{\gamma} u_2$$

$$B. y = \frac{\alpha}{\gamma}(\mathrm{u}_1 + \mathrm{u}_2) \rightarrow u_1 = 0 \rightarrow \frac{\alpha}{\gamma}u_2$$

כפי שניתן לראות, מאחר ו $u_1 = 0$ אין הבדל בין PCA שמבצע הורדת מימדים מ2 ל1 ולרשת נוירונים הלינארית מאחר ואין איבוד של מידע מאחר והוא שווה ל0.

מקרה 2:

$$(ru_1) = \alpha_1 = \alpha, (ru_2) = \alpha_2 = \alpha$$
$$(u_1^2) = \gamma_1 = \gamma\,, (u_2^2) = \gamma_2 = \gamma$$
$$(u_1 u_2) = \beta > 0$$

$$C = \begin{pmatrix} \gamma_1 & \beta \\ \beta & \gamma_2 \end{pmatrix} = \begin{pmatrix} \gamma & \beta \\ \beta & \gamma \end{pmatrix}$$
$$|C| = \gamma^2 - \beta^2 = (\gamma - \beta) + (\gamma + \beta) \rightarrow \gamma = \pm\beta \rightarrow \lambda_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \lambda_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

PCA:

$$y = wz$$

$$w = \eta\frac{rz}{z^2}$$

$$\frac{rz}{z^2} = \frac{r(j_1u_1 + j_2u_2)}{(j_1u_1 + j_2u_2)^2} \rightarrow j_1 = \frac{1}{\sqrt{2}}, j_2 = \frac{1}{\sqrt{2}} \rightarrow \frac{\frac{1}{\sqrt{2}}ru_1 + \frac{1}{\sqrt{2}}ru_2}{u_1^2 + 2u_1u_2 + u_2^2} = \frac{\frac{2}{\sqrt{2}}\alpha}{2\gamma + 2\beta} = \frac{\sqrt{2}\alpha}{\gamma + \beta}$$

$$y = wz = w(j_1u_1 + j_2u_2) \rightarrow j_1 = \frac{1}{\sqrt{2}}, j_2 = \frac{1}{\sqrt{2}} \rightarrow= w\left(\frac{1}{\sqrt{2}}\mathrm{u}_1 + \frac{1}{\sqrt{2}}\mathrm{u}_2\right) = \frac{\alpha}{\gamma + \beta}\cdot\sqrt{2}(\mathrm{u}_1 + \mathrm{u}_2)$$

Linear network:

$$w_2 = \frac{\alpha}{\beta + \gamma}$$

$$w_1 = \frac{\alpha}{\beta + \gamma}$$

$$y = w_1u_1 + w_2u_2 = (\mathrm{u}_1 + \mathrm{u}_2)\frac{\alpha}{\beta + \gamma}$$

$$\frac{\mathrm{y}_{\mathrm{linear}}}{y_{pca}} = \frac{(\mathrm{u}_1 + \mathrm{u}_2)\frac{\alpha}{\beta + \gamma}}{\frac{\alpha}{\gamma + \beta}\cdot\sqrt{2}(\mathrm{u}_1 + \mathrm{u}_2)} = \frac{1}{\sqrt{2}}$$

כפי שניתן לראות גם במקרה זה ישנו דמיון בין שתי הדרכים. בשלב הPCA עקב הורדת הממדים ה"גסה" שמתבצעת בדרך זו, משניים לאחד ישנה טרנספורמציה לינארית לממוצע בהכפלה בשורש 2. אי לכך התוצאה תסטה עד לפאקטור של שורש 2 = 1.41 מן תוצאות הרשת הלינארית, שנוכל לתאר אותן כמדויקות יותר.

2. א.

ו)

ראשית, ניזכר בזהות התפלגות גאוסית שתשרת אותנו:

$$E(X^2) = Var(x) + [E(X)^2]$$

ולכן נציין לפי הנתונים ולפי הזהות כי:

$$\langle S\rangle = \langle Z\rangle = 0$$
$$\langle S^2\rangle = \sigma_s^2 + 0 = \sigma_s^2$$
$$\langle Z^2\rangle = \sigma_h^2 + 0 = \sigma_h^2$$

$$\epsilon = \frac{1}{2}(S-Y)^2 = \frac{1}{2}\left(S - W(S+Z)\right)^2 = \frac{1}{2}(S^2 - 2SW(S+Z) + W^2S^2 + 2W^2SZ + W^2Z^2)$$
$$= \frac{1}{2}(S^2 - 2S^2W - 2SWZ + W^2S^2 + 2W^2SZ + W^2Z^2)$$
$$= \frac{1}{2}\left(S^2 - SW(S^2 + SZ) + W^2(S^2 + 2SZ + Z^2)\right)$$

כעת נמצע תוך התחשבות בכך ש- S ו- Z בלתי תלויים:

$$\frac{1}{2}(\langle S^2\rangle - 2W(\langle S^2\rangle + \langle S\rangle\langle Z\rangle) + W^2(\langle S^2\rangle + 2\langle S\rangle\langle Z\rangle + \langle Z^2\rangle)$$

נתון כי התוחלת של S ו-Z = 0:

$$\frac{1}{2}(\langle S^2\rangle - 2W(\langle S^2\rangle) + W^2(\langle S^2\rangle + \langle Z^2\rangle) =$$
$$\frac{1}{2}(\langle S^2\rangle - 2W\langle S^2\rangle + W^2(\langle S^2\rangle + \langle Z^2\rangle) =$$
$$\frac{1}{2}\left(\sigma_s^2 - 2W(\sigma_s^2) + W^2\left(\sigma_s^2 + \sigma_h^2\right)\right) =$$
$$\frac{1}{2}\left(\sigma_s^2 - 2W\sigma_s^2 + W^2\left(\sigma_s^2 + \sigma_h^2\right)\right) =$$

כעת נגזור לפי W:

$$\epsilon'(W) = \frac{1}{2}\left(-2\sigma_s^2 + 2W\left(\sigma_s^2 + \sigma_h^2\right)\right) =$$
$$\frac{1}{2}(2\left(-\sigma_s^2 + W\left(\sigma_s^2 + \sigma_h^2\right)\right) =$$
$$\epsilon'(W) = \left(-\sigma_s^2 + W\left(\sigma_s^2 + \sigma_h^2\right)\right) =$$

כעת נשווה ל-0 למציאות מינימום:

$$-\sigma_s^2 + W\left(\sigma_s^2 + \sigma_h^2\right) = 0$$
$$W\left(\sigma_s^2 + \sigma_h^2\right) = \sigma_s^2$$

$$W = \frac{\sigma_s^2}{(\sigma_s^2 + \sigma_h^2)}$$

(ii

נתון כי $Y = W(S + Z)$

נמצע לפי Z:

$$Y = W(S + \langle Z \rangle) = W(S + 0) = WS$$
$$Y = \frac{\sigma_s^2 * S}{(\sigma_s^2 + \sigma_h^2)}$$

לטובת השאלה הבאה, נכתוב המשוואה כך:

$$Y = \frac{1}{1 + \left(\frac{\sigma_h^2}{\sigma_s^2}\right)} * S$$

כעת, עבור $\sigma_s^2 \gg \sigma_h^2$:

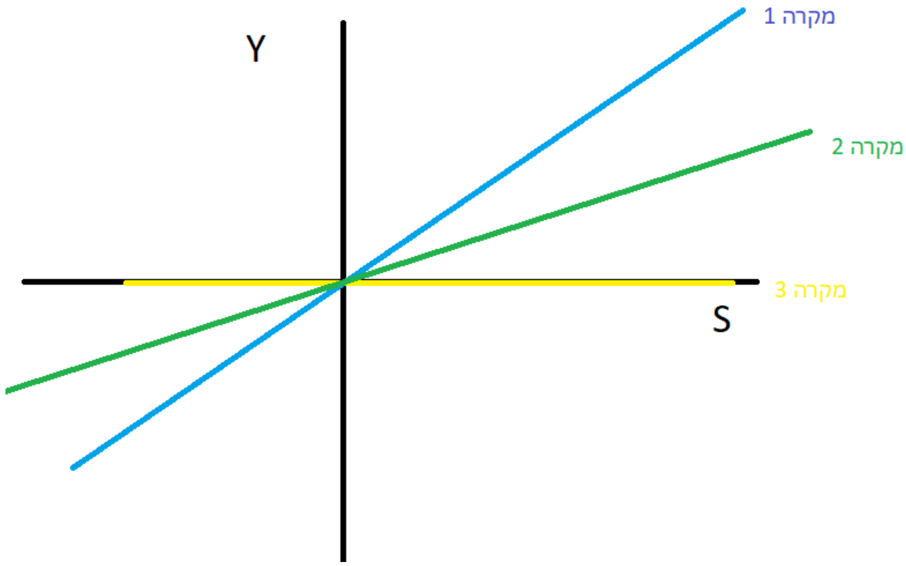
$$\lim Y = \frac{1}{1+0} * S = S$$

עבור $\sigma_s^2 = \sigma_h^2$:

$$\lim Y = \frac{1}{1+1} * S = \frac{S}{2}$$

עבור $\sigma_s^2 \ll \sigma_h^2$:

$$\lim Y = \frac{1}{1+\infty} * S = 0$$



ב)

נתון:

$$\vec{U} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\vec{X} = \begin{pmatrix} U_1 S + Z_1 \\ U_2 S + Z_2 \end{pmatrix} = \begin{pmatrix} S + Z_1 \\ S + Z_2 \end{pmatrix}$$

לכן:

$$Y = \vec{W}\vec{X} = (W_1 \;\; W_2)\begin{pmatrix} S + Z_1 \\ S + Z_2 \end{pmatrix} = W_1(S + Z_1) + W_2(S + Z_2)$$

שגיאה ריבועית:

$$\epsilon = \frac{1}{2}(S - Y)^2 = \frac{1}{2}\left(S - \left(W_1(S + Z_1) + W_2(S + Z_2)\right)\right)^2 =$$

$$\frac{1}{2}\Big(S^2 - 2S\big(W_1(S + Z_1) + W_2(S + Z_2)\big) + W_1^2(S^2 + 2SZ_1 + Z_1^2) + 2W_1(S + Z_1)W_2(S + Z_2) + W_2^2(S^2 + 2SZ_2 + Z_2^2)\Big) =$$

$$\frac{1}{2}\Big(S^2 - 2W_1(S^2 + Z_1 S) - 2W_2(S^2 + SZ_2) + W_1^2(S^2 + 2SZ_1 + Z_1^2) + 2W_1W_2(S^2 + SZ_2 + SZ_1 + Z_1Z_2) + W_2^2(S^2 + 2SZ_2 + Z_2^2)\Big) =$$

כעת נמצע לפי S ו-Z:

$$\frac{1}{2}(\langle S^2\rangle - 2W_1(\langle S^2\rangle + \langle Z_1\rangle\langle S\rangle - 2W_2(\langle S\rangle + \langle Z_2\rangle) + W_1^2(\langle S^2\rangle + 2\langle S\rangle\langle Z_1\rangle + \langle Z_1^2\rangle) + 2W_1W_2(\langle S^2\rangle + \langle S\rangle\langle Z_2\rangle + \langle S\rangle\langle Z_1\rangle + \langle Z_1\rangle\langle Z_2\rangle) + W_2^2\big(\langle S^2\rangle + 2\langle S\rangle\langle Z_2\rangle + \langle Z_2^2\rangle\big) =$$

$$\frac{1}{2}(\sigma_s^2 - 2W_1\sigma_s^2 - 2W_2\sigma_s^2 + W_1^2(\sigma_s^2 + \sigma_h^2) + 2W_1W_2\sigma_s^2 + W_2^2(\sigma_s^2 + \sigma_h^2) =$$

כעת נגזור את שני ה-Wים ונשווה ל-0:

$$\frac{d\epsilon}{dW_1}=\frac{1}{2}\left(-2\sigma_s^2+2W_1\left(\sigma_s^2+\sigma_h^2\right)+2W_2\sigma_s^2\right)=W_1\left(\sigma_s^2+\sigma_h^2\right)+W_2\sigma_s^2-\sigma_s^2=0$$
$$\frac{d\epsilon}{dW_2}=\frac{1}{2}\left(-2\sigma_s^2+2W_1\sigma_s^2+2W_2\left(\sigma_s^2+\sigma_h^2\right)\right)=W_1\sigma_s^2+W_2\left(\sigma_s^2+\sigma_h^2\right)-\sigma_s^2=0$$

כעת נשווה בין שתי המשוואות, הרי שתיהן שוות ל-0:

$$\begin{aligned}
&W_1\left(\sigma_s^2+\sigma_h^2\right)+W_2\sigma_s^2-\sigma_s^2=W_1\sigma_s^2+W_2\left(\sigma_s^2+\sigma_h^2\right)-\sigma_s^2\\
&W_1\left(\sigma_s^2+\sigma_h^2\right)+W_2\sigma_s^2=W_1\sigma_s^2+W_2\left(\sigma_s^2+\sigma_h^2\right)\\
&W_1\left(\sigma_s^2+\sigma_h^2\right)-W_1\sigma_s^2=W_2\left(\sigma_s^2+\sigma_h^2\right)-W_2\sigma_s^2\\
&W_1\sigma_s^2+W_1\sigma_h^2-W_1\sigma_s^2=W_2\sigma_s^2+W_2\sigma_h^2-W_2\sigma_s^2\\
&W_1\sigma_h^2=W_2\sigma_h^2\\
&W_1=W_2
\end{aligned}$$

בהצבת התנאים הנתונים בשאלה, מתקבל כי עבור 2 מימדים מתקבל כי $W_1=W_2$. לכן, עבור $\overrightarrow{W}$ כללי, מתקיים:

$$\frac{d\epsilon}{dW}=W\left(\sigma_s^2+\sigma_h^2\right)+W\sigma_s^2-\sigma_s^2$$

וכמשווים את המשוואה הנ"ל ל-0 מתקבל:

$$\begin{aligned}
&W\left(\sigma_2^2+\sigma_h^2\right)+W\sigma_s^2-\sigma_s^2=0\\
&W\left(\sigma_2^2+\sigma_h^2\right)+W\sigma_s^2=\sigma_s^2\\
&W\left(\left(\sigma_2^2+\sigma_h^2\right)+\sigma_s^2\right)=\sigma_s^2\\
&W=\frac{\sigma_s^2}{\left(\sigma_2^2+\sigma_h^2\right)+\sigma_s^2}\\
&W=\frac{\sigma_s^2}{2\sigma_s^2+\sigma_h^2}
\end{aligned}$$

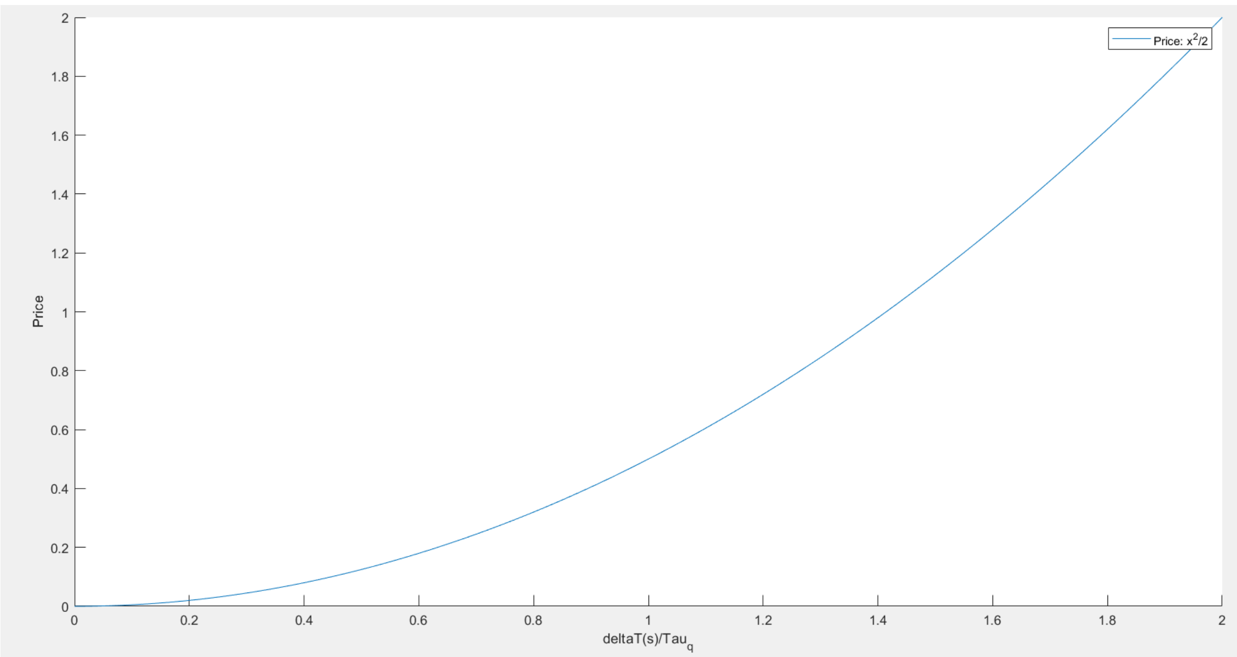
ג.

לפי הסעיפים הקודמים, תחת האילוצים הנתונים ($U_1=U_2=1$), הראינו כי ה-W האופטימלי עבור N מימדים שבדקנו (1,2) התשובה היתה $W=\frac{\sigma_s^2}{N\sigma_s^2+\sigma_h^2}$. דרך התבוננות בחישובים אלו והתפתחותם, ניתן להניח כי עבור $U_1=\cdots=U_M=1$ ועבור $W_1=\cdots W_M$ יתקיים שה-W האופטימלי עבור M נוירוני ביניים הינו:

$$W=\frac{\sigma_s^2}{M\sigma_s^2+\sigma_h^2}$$

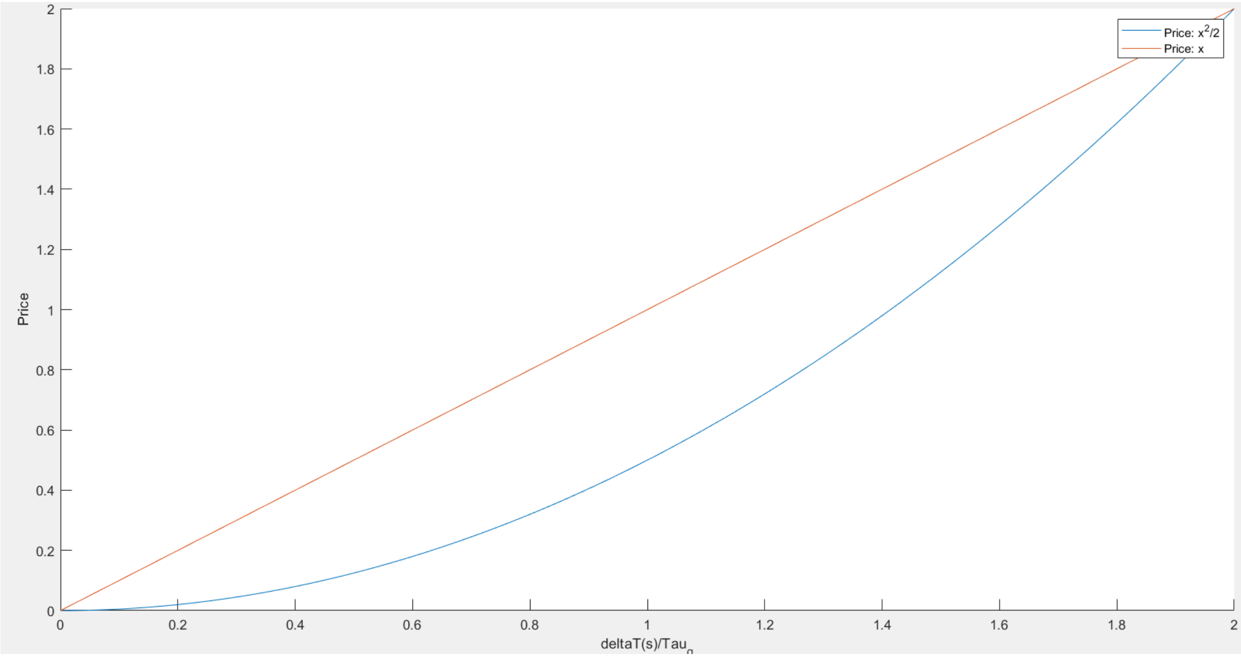
<u>חלק 2</u>

1. המחיר על הזזת הספייק משנה את דרך הלמידה של הכרונוטרון. פונקציית סיגמא קובעת אילו ספייקים נרצה להסיר, ליצור את להזיז בהתאם למחיר הנקבע מהספייק. כלומר, ככל שהספייק רחוק יותר מהספייק של המורה, כך המחיר שלו גבוהה יותר. כמה גבוה יותר? בהתאם לפונקציית סיגמא הנבחרת. המחיר עבור סיגמא הינו $\sigma = \left(\frac{\left|t_i^j - \hat{t}_i^k\right|}{\tau_q}\right)$ כאשר t_i^j הינו הספייק הנלמד (ברגע זה, נתון לשינוי.) $\hat{t}_i^k$ הינו הספייק של המורה. ההפרש בניהם משקל את deltaT בפאקטור של τ_q שהינו קבוע. ידוע מנתוני השאלה כי $\max\left(\left|t_i^j - \hat{t}_i^k\right|\right) = 0.5$ נניח לצורך ההדגמה כי $\tau_q = 0.25$
נוכל להתבונן בפונקציה הריבועית שנבחרה: $\sigma = \frac{x^2}{2}$



השיפוע בפונציה זו עולה ככל שהפרש במרחק גדל. ולכן הפונקציה תיהיה "סלחנית" יותר עבור ספייקים שנלמדו קרוב לזמן המדוייק שהמורה ציין. ו"מחמירה" יותר עבור ספייקים רחוקים מהזמן המדוייק שהמורה ציין.

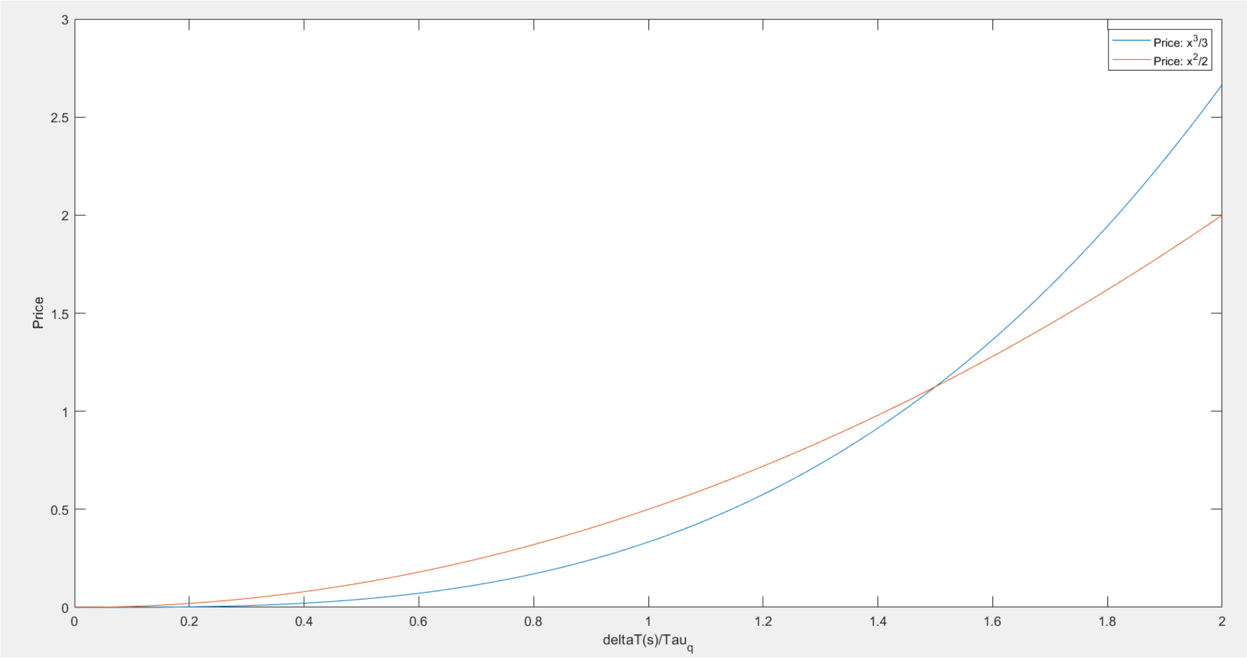
נשווה זאת לפונקציה עולה לינארית $\sigma = x$



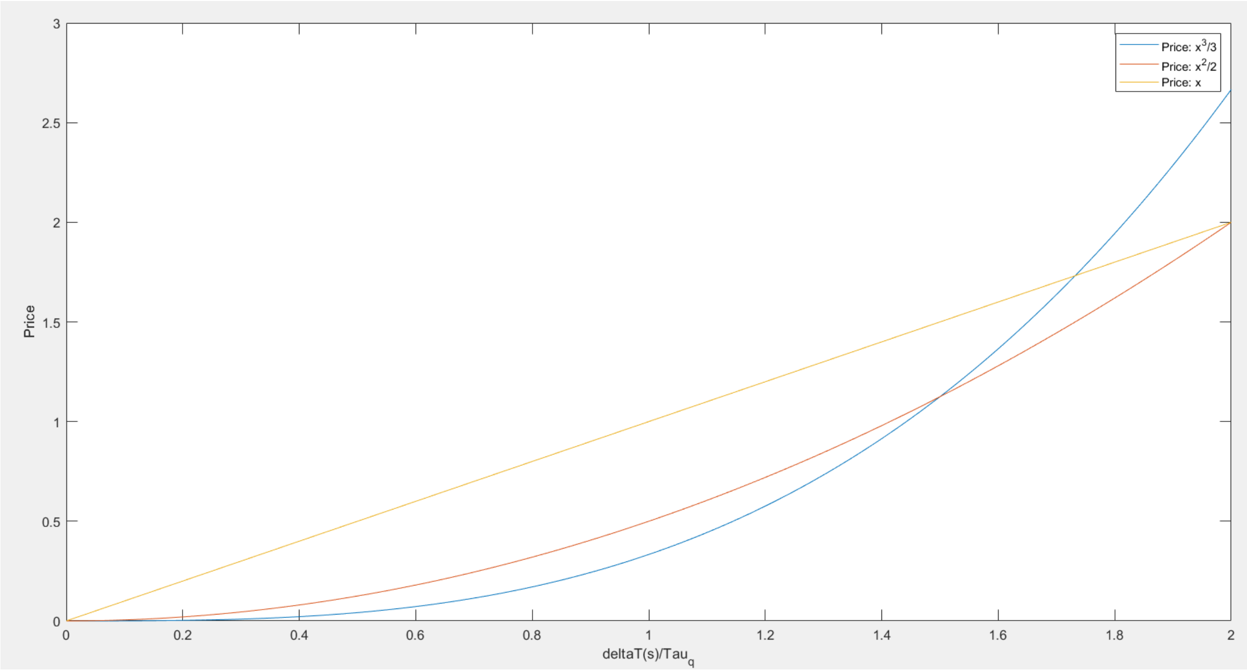
בפונקציה זו השיפוע הינו קבוע ולכן היא אינה מחמירה את סלחנית עם ספייקים קרובים או רחוקים. פונקציה זו תגיע להצלחה בלמידה לאחר מספר גדול יותר של ניסיונות בממוצע מאחר הפונקציה הפרבולית במקרים בהם תוכל ללמוד, היא תדרוש מספר אפוקים גדול יותר. הלמידה במקרה זה הינה איטית יותר מהפונקציה הפרבולית הנדונה. הבעיה בפונקציה זו הינה חוסר הקירבה ל0 אלא רק ב0 עצמה. מצב שיווה בעיה ללמידה מרובת ספייקים בה למורה מספק ספייקים שדרושים להילמד. במצב זה אלא אם הספייקים של המורה באותו הזמן, יהיה קושי רב ללמידה מוצלחת. אם במצב זה נניח אפוק אינסוף, ותנאי יציאה להתאמה בין הספייקים הנלמדים למורה, הכרונוטרון ינסה ללמוד עד אינסוף ולעולם לא יצליח.

לסיכום, נצפה כי הפונ' הפרבולית תלמד מהר יותר לאחר פחות איפוקים ובקירוב גדול יותר. ניתן לראות כי לאחר *deltaT/tau_q=2* הפונ' הפרבולית תיהיה מחמירה יותר ותגבה מחיר גבוה יותר עבור טעות כה חמורה בין הכרונוטרון למורה.

במקרה של $\sigma = \frac{x^3}{3}$



פונקצית המחיר תיהיה סלחנית יותר עבור הפרש זמנים קטנים. מאחר ופונ' המחיר של $\frac{x^3}{3}$ עולה מעל $\frac{x^2}{2}$ ב*deltaT/tau_q=1.5* לאחר נק' זו נראה כי פונ' זו בעלת מחיר גבוהה יותר ולכן מחמירה יותר עבור הפרשי זמנים גדולים. הרי כי פונ' $\frac{x^2}{2}$ תלמד לאט יותר אבל בטוח יותר במקרה של למידת הזמנים הנתונה. ידרשו יותר איפוקים על מנת להגיע למשקלים האידיאלים אך סיכויי ההצלחה הינם גבוהים יותר כי לא תיהיה קפיצה מעל ערכים אלו עקב תשלום מחיר גבוה מידי.



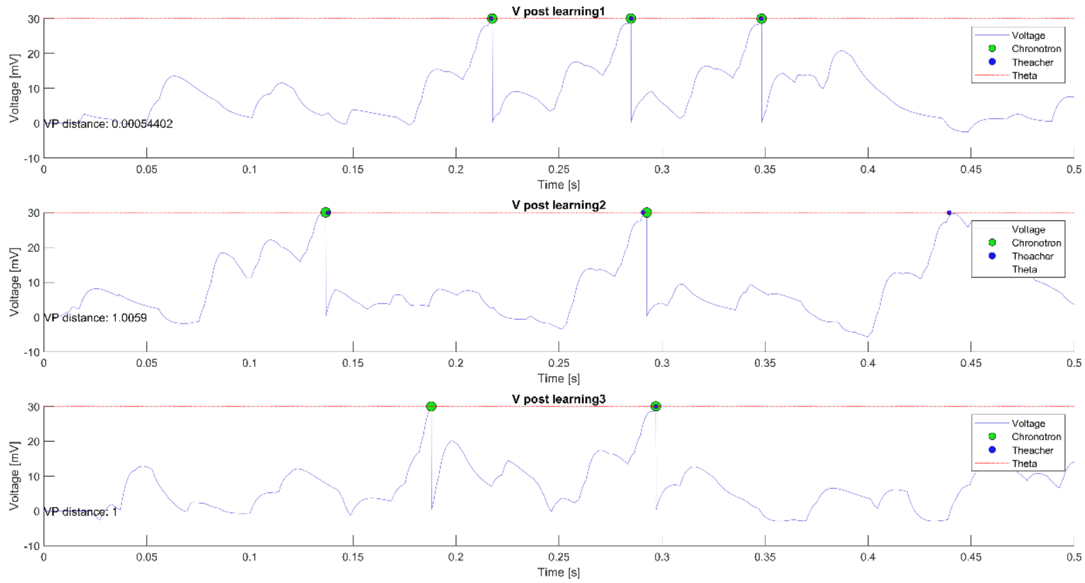
לסיכום,

פונ' x היא אינה סלחנית ואינה מחמירה.

פונ' $x^2/2$ הינה סלחנית יותר מx ומחמירה יותר.

פונ' $x^3/3$ הינה הסלחנית והמחמירה ביותר מבין הפונקציות.

.2



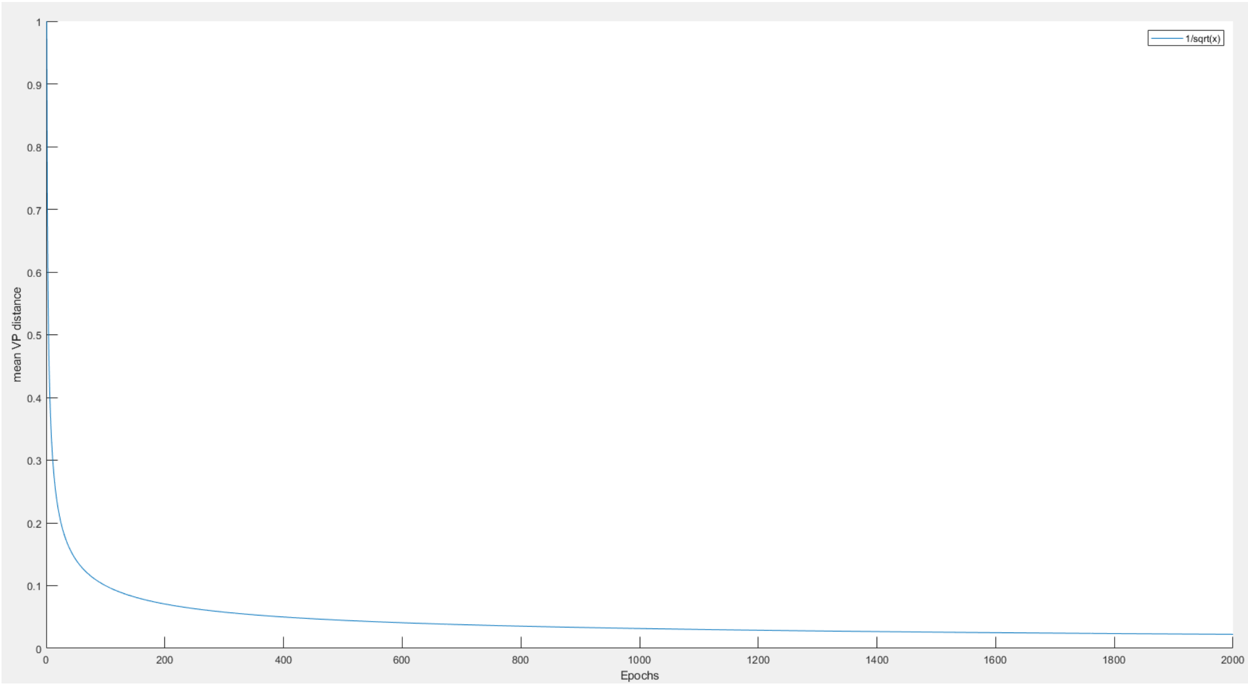
Mean distance: 0.213489

Standtard deviation: 0.470671

Learning time: 8.31

3. מרחק Victor-Purpura הממוצע מבטא את הצלחת הלמידה של הכרונוטרון על פני כל הדוגמאות שניתנו. המרחק Victor-Purpura עבור דוגמא נתונה מבטא את הצלחת הלמידה של הכרונוטרון בדוגמא הספציפית הזאת. כך שאם הזמן בו הכרונוטרון טוען כי קיים ספייק שווה לזמן בו המורה אומר שיש ספייק המרחק VP=0. ממוצע כל המרחקים של VP על פני כל הדוגמאות מבטא את הצלחת הכרונוטרון בלמידת כל הדוגמאות. המרחק הממוצע מבטא הצלחה יחסית, הרי שככל שיש יותר דוגמאות נתונות כך הלמידה יותר "קשה" ולכן ההצלחה נמדדת באופן שיוויוני בעזרת חילוק במספר הדוגמאות – כחלק מהגדרת המושג ממוצע.
בדוגמא שלנו הכרונוטרון המרחק הממוצע של הכרונוטרון קטן מ1. לאחר ריצה של 200 אפוקים, 200 פעם לרוץ על כל הדוגמאות הנתונות יתקבלו תוצאות של 0.26 בקירוב. בעוד ריצה של 1800 אפוקים מגיעה ל0.24 בקירוב. כך שניתן לראות כי קיים סף למרחק המינימלי. ייתכן כי באינסוף אפוקים יגיע המרחק ל0. אך בכל מספר רציונלי של אפוקים שנבחר מעל 100, נקבל תוצאה בין 0 ל0.7 בקירוב.

4. הכרונוטרון למד בהצלחה את הדוגמאות הנתונות, ניתן לראות כי עבור הדוגמאות הנתונות באופן קונסיסטנטי ככל שגדלים מספר האפוקים כך יורד מרחק VP. משמע, ככל שניתן לכרונוטרון יותר זמן ונתונים לרוץ, הוא ילמד טוב יותר את כל הדוגמאות ויגיע לערכי משקלים אופטימלים עבור יותר דוגמאות. הכרונוטרון ידייק באופן מושלם (בקירוב של עשרות אלפי השניה) רק באינסוף, ובמצב זה מרחק VP יהיה 0.
ניתן להשוות את פונקציית הלמידה של הכרונוטרון ל$y = \frac{1}{\sqrt{x}}$



כאשר כפי שניתן לראות ככל שעולים מספר האיפוקים ערך המרחק *VP* קטן, אך גם השינוי (שיפוע) של של המרחק קטן ככל שמס' האיפוקים עולה.
נצפה שבאינסוף יגיע ל0, ולכן הלמידה הצליחה.