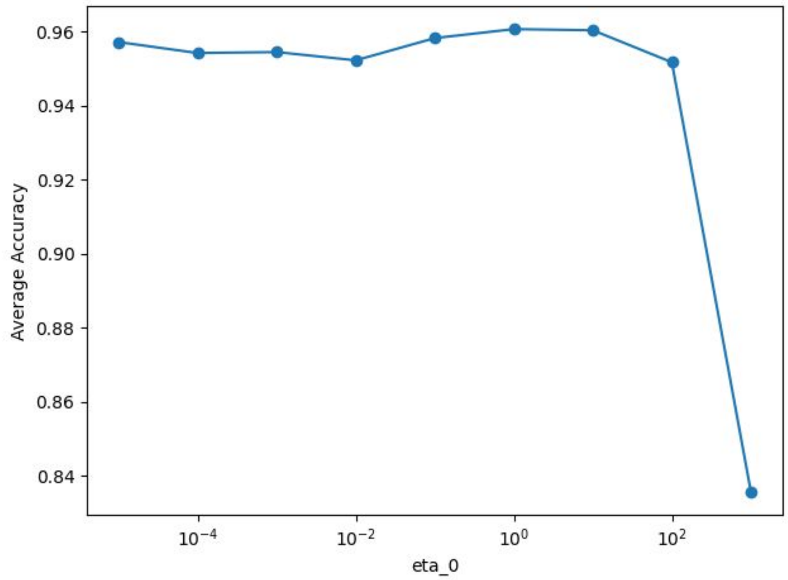
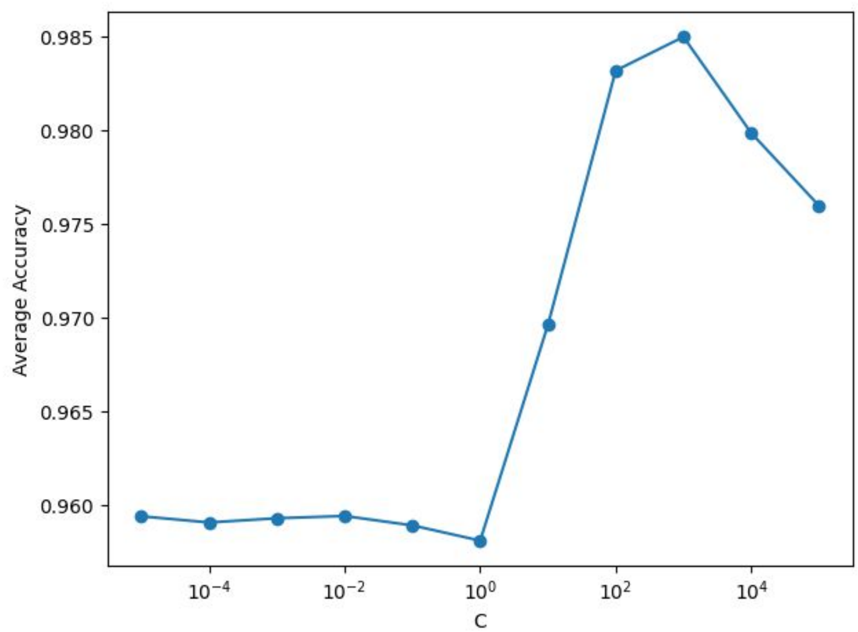
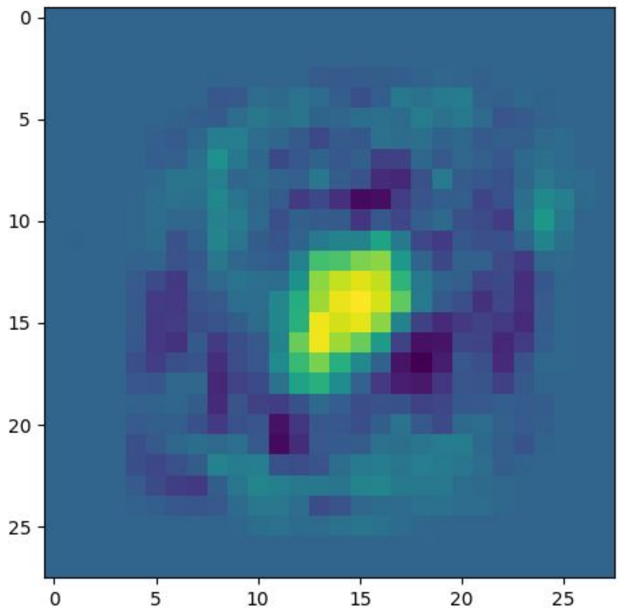
<u>Question 1a</u> - accuracy per eta_0 (best accuracy is for eta_0=1):



<u>Question 1b</u> - accuracy per C (best accuracy is for c=1000):

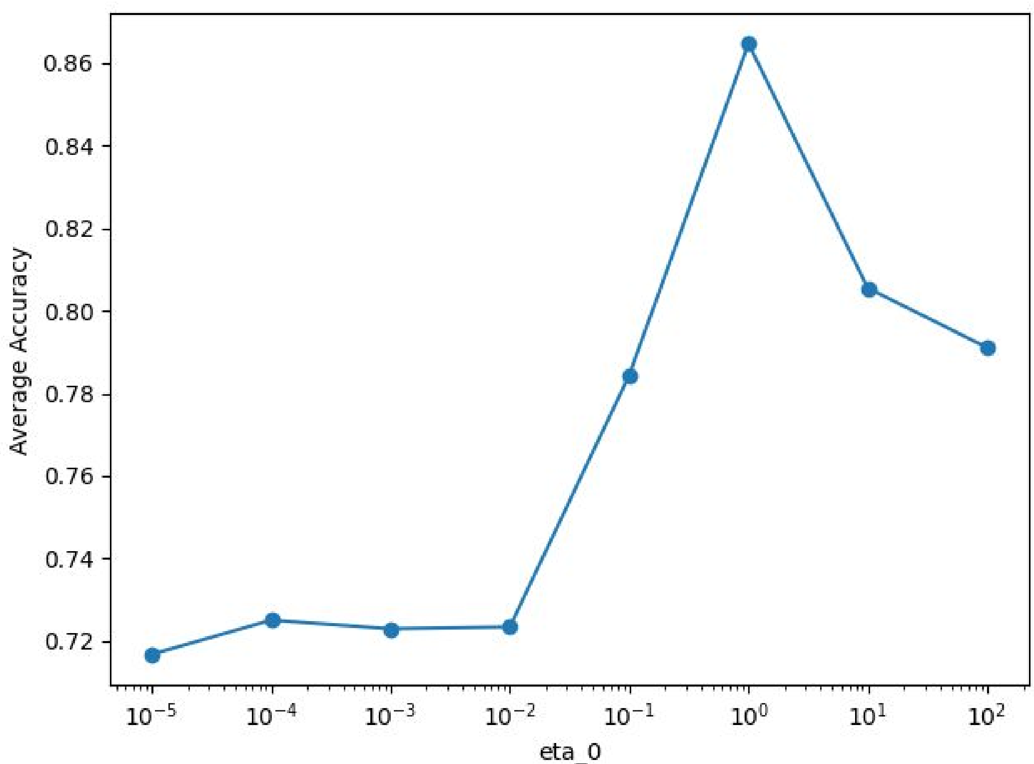


Question 1c - the SGD result

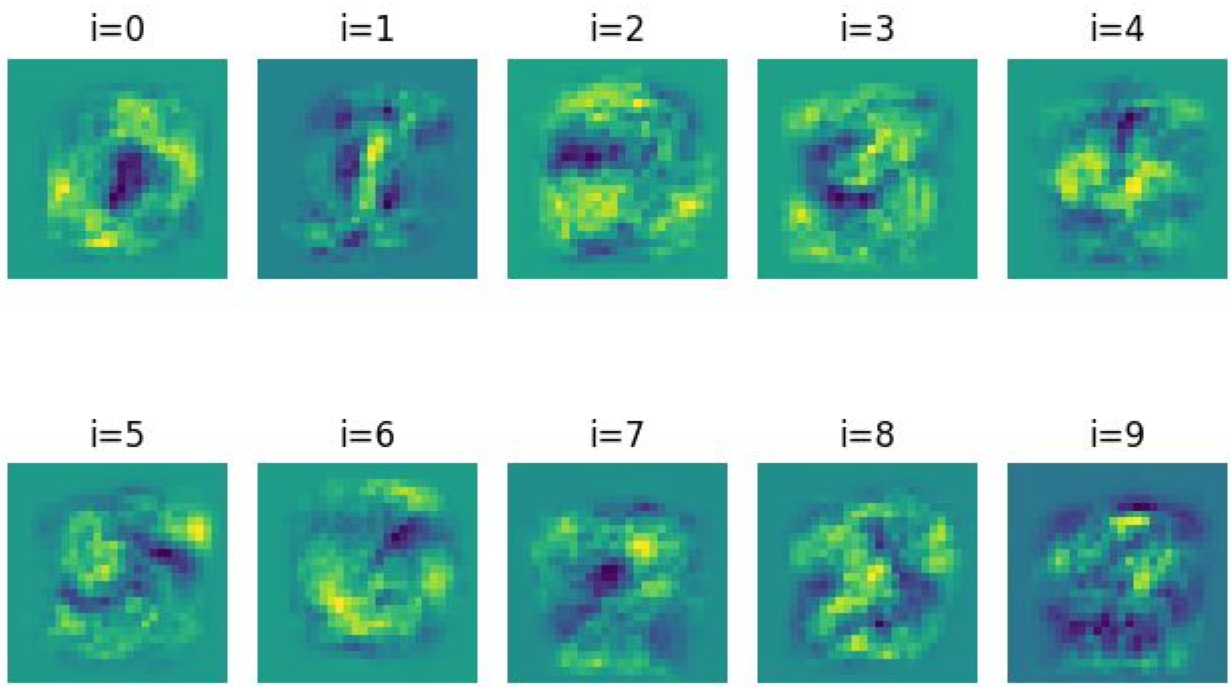


Question 1d - the final accuracy on the test data is **0.9918116683725691**

Question 2a - accuracy per eta_0 (best accuracy is for eta_0=1):



Question 2b



Question 3c - the final accuracy on the test_data is **0.8885**

מבוא ללמידה חישובית - תרגיל בית 4

ד) יהיו נקודות x,y ו-$\alpha\in[0,1]$. צריך להראות ש-
$g(\alpha x+(1-\alpha)y)\le \alpha g(x)+(1-\alpha)g(y)$. נניח באינדוקציה על m.
בסיס - $m=1 \Leftarrow g=f_1$ וסיימנו.
הנחה: נניח ש- $h=\max\limits_{1\le i\le m-1} f_i$ קמורה (נניח עבורו) ויהי f_m פונקציה קמורה. צעד: נסמן $z=\alpha x+(1-\alpha)y$ ונקבל כעת:

$$g(z)=\max(f_m(z),h(z))\le$$
$$\le \max(\alpha f_m(x)+(1-\alpha)f_m(y),\ \alpha h(x)+(1-\alpha)h(y))\le$$
$$\le \max(\alpha[f_m(x)+h(x)],\ (1-\alpha)[f_m(y)+h(y)])=$$
$$=\alpha\max(f_m(x),h(x))+(1-\alpha)\max(f_m(y)+h(y))=$$
$$=\alpha g(x)+(1-\alpha)g(y)$$

ה) יהי $X\in\mathcal{X}^m$. מספיק להראות ש- $|F_X|\le \pi_{F_1}(m)\cdot\pi_{F_2}(m)$ מכיוון ש-X שרירותי.

$$|F_X|=|\{f_2(f_1(x_1)),\dots,f_2(f_1(x_m))\mid f_1\in F_1,f_2\in F_2\}|=$$
$$=\Big|\bigcup_{y\in F_{1X}}\{f_2(y_1),\dots,f_2(y_m))\mid f_2\in F_2\}$$

מחסם האיחוד (Union Bound) נקבל:

$$\le\sum_{y\in F_{1X}}|\{f_2(y_1),\dots,f_2(y_m)\}_{f_2\in F_2}|\le\sum_{y\in F_{1X}}\pi_{F_2}(m)\le\pi_{F_1}(m)\cdot\pi_{F_2}(m)$$

כנדרש.

הערה: הסימון F_X או F_{1X} שקול לסימון H_C בהרצאה - כל התיוגים האפשריים של פונקציות מ-F על המקבץ X.

5) נתון: $l(x_{t+1}) \le l(x_t) + \nabla l(x_t)^t (x_{t+1}-x_t) + \frac{\beta}{2}\|x_t - x_{t+1}\|^2$

נציב ש- $x_{t+1}-x_t = -\eta \nabla l(x_t)$ ונקבל:

$$l(x_{t+1}) \le l(x_t) + \nabla l(x_t)^t(-\eta \nabla l(x_t)) + \frac{\beta}{2}\|\eta \nabla l(x_t)\|^2 =$$

$$= l(x_t) - \eta \nabla l(x_t)^t \nabla l(x_t) + \frac{\beta}{2}\eta^2 \|\nabla l(x_t)\|^2$$

$$\circledast = l(x_t) + \left(\frac{\beta}{2}\eta^2 - \eta\right)\|\nabla l(x_t)\|^2$$

כי $x^t x = \|x\|^2$

ואז $\eta < \frac{2}{\beta}$ ולכן: $\frac{\beta}{2}\eta^2 - \eta < \frac{\beta}{2}\cdot\frac{4}{\beta^2} - \frac{2}{\beta} = \frac{2}{\beta} - \frac{2}{\beta} = 0$

ולכן כשנחלק ב- $\frac{\beta}{2}\eta^2 - \eta$ הסימן מתהפך ונקבל:

$$\|\nabla l(x_t)\|^2 \le \frac{l(x_{t+1}) - l(x_t)}{\frac{\beta}{2}\eta^2 - \eta}$$

ולכן:

$$\sum_{t=0}^{n} \|\nabla l(x_t)\|^2 \le \sum_{t=0}^{n} \frac{l(x_{t+1}) - l(x_t)}{\frac{\beta}{2}\eta^2 - \eta} = \frac{1}{\frac{\beta}{2}\eta^2 - \eta} \sum_{t=0}^{n} l(x_{t+1}) - l(x_t) =$$

וכעת זה סכום טלסקופי ולכן:

$$= \frac{-1}{\frac{\beta}{2}\eta^2 - \eta}\left[l(x_0) - l(x_{n+1})\right] \le \frac{1}{\eta - \frac{\beta}{2}\eta^2} l(x_0) := K < \infty$$

כאשר המעבר האחרון נובע מכך ש- $l(x_{n+1}) \ge 0$

קיבלנו - $\sum_{t=0}^{n} \|\nabla l(x_t)\|^2 < K$ כאשר K לא תלוי ב-n

אלא רק ב- $\eta, \beta, l(x_0)$ לכן זה נכון לכל n,

והטור חסום ע"י קבוע לכל n ומתכנס כנדרש.

5. לפי שאלה 1 מספיק להוכיח שלכל $\hat{y}\in L$ - פונקציה

כלשהו מתקיים "φ": $\ell(w_1,\dots,w_L,x,y)=w_{\hat{y}}\cdot x-w_y\cdot x+\Delta_{zo}(\hat{y},y)$ פונקציה

קמורה וסיימנו כי מקסימום של פונקציות קמורות הוא

קמור.

במקום הזה שצירוף לינארי של פונקציות

קמורות בפ הוא קמור, לכן מספיק להראות במקום

ל- $\Delta_{zo}(\hat{y},y)$, $w_y\cdot x$, $w_{\hat{y}}\cdot x$ קמורות.

- $\Delta_{zo}(\hat{y},y)$ - טריוויאלי (קבוע) של פונקציות קבועות)
- $w_{\hat{y}}\cdot x, w_y\cdot x$ פונקציות לינאריות ביחס ל- $w_y / w_{\hat{y}}$ לכן גם קמורות כנדרש.

ג. יהיו y, x, f. נסמן ב- y^* את התווית ש- f מסווג ל- x, כלומר $f(x;w_1,\dots,w_L)=y^*$. אז:

$$\Delta_{zo}(f(x;w_1,\dots,w_L),y)=\Delta_{zo}(y^*,y)\le$$

$$\circledast\le\Delta_{zo}(y^*,y)+w_{y^*}x-w_y\cdot x\le$$

$$\le\max_{\hat{y}\in[L]}\left(\Delta_{zo}(\hat{y},y)+w_{\hat{y}}\cdot x-w_y\cdot x\right)=\ell(w_1,\dots,w_L,x,y)$$

כנדרש.

המעבר $\circledast$ נובע מכך ש- $w_{y^*}\cdot x\ge w\cdot x$ - ש מאחר שהגדרנו את y^*,

$w_{y^*}x\ge w_y x$ לכל $y\in[L]$.

3) נניח בשלילה שיש j כך: $\hat{y} := f(x_j; w_1^{opt},\dots,w_L^{opt}) \neq y_j$

כי הניבוי לא מדויק. אז:

$$\ell(w_1^{opt},\dots,w_L^{opt},x_j,y_j) = \underbrace{\Delta_{zo}(\hat{y},y_j)}_{=1} + \underbrace{w^{opt}_{\hat{y}}x_j - w^{opt}_{y_j}x_j}_{\geq 0 \text{ לפי הגדרת } f \text{ מהמחה}} \geq 1$$

באופן דומה ניתן לראות שעבור w^* מתקיים לכל $\hat{y}, i$:

$$\ell(w_1^*,\dots,w_L^*,x_i,y_i) \leq \Delta_{zo}(\hat{y},y_i) + w^*_{\hat{y}}x_i - w^*_{y_i}x_i$$

$$= \begin{cases} 0, & \hat{y}=y_i \\ 1+w^*_{\hat{y}}x_i - w^*_{y_i}x_i \end{cases} \leq 1$$

כי $w^*_{\hat{y}}x_i \leq w^*_{y_i}x_i$ לכל $i, \hat{y}$ לפי הגדרת w^*.

כלומר קיבלנו דוגמה (x_j,y_j) שבה w^* מקבל Loss קטן יותר מ-w^{opt}. אבל נתון:

$$\sum_i \ell(w_1^{opt},\dots,w_L^{opt},x_i,y_i) \leq \sum_i \ell(w_1^*,\dots,w_L^*,x_i,y_i)$$

לכן יש j כך ש- $\ell(w_1^*,\dots,w_L^*,x_j,y_j) > \ell(w_1^{opt},\dots,w_L^{opt},x_j,y_j)$ וגם בה"כ $\|x_i\|=1$ לכל i

נניח בה"כ ש- $\forall k: \|w_k^{opt}\| = \|w_k^*\| = 1$. מהשורה הקודמת נקבל:

$$\exists \hat{y}\,\forall \bar{y}.\ w^*_{\hat{y}}x_j - w^*_{y_j}x_j + \Delta_{zo}(\hat{y},y_j) > w^{opt}_{\bar{y}}x_j - w^{opt}_{y_j}x_j + \Delta_{zo}(\bar{y},y_j)$$

אם $\hat{y}=y_j$ אז אגף שמאל מתאפס אבל אגף ימין אי-שלילי,

לכן $\hat{y}\neq y_j$, $\Delta_{zo}(\hat{y},y_j)=1$.

בנוסף הנ"ל נכון לכל $\bar{y}$ בפרט $\bar{y}=\hat{y}$:

$$w^*_{\hat{y}}x_j - w^*_{y_j}x_j + 1 > w^{opt}_{\hat{y}}x_j - w^{opt}_{y_j}x_j + 1$$

$$\Downarrow$$

$$(w^*_{\hat{y}} - w^*_{y_j})x_j > (w^{opt}_{\hat{y}} - w^{opt}_{y_j})x_j$$

אם נמצא סתירה כי $w^*_{\hat{y}} - w^*_{y_j}$ יכול להיות קטן כרצוננו שכן אם נכפיל את w_k ב-$\alpha>0$ כלשהו לכל $k\in[L]$ הגדרת w^* לא תשתנה - הוא עדיין לא יטעה, כלומר יש w^* חוקי עבורו הא"ש הנ"ל מתקיים לכל $x_j, y_j, \hat{y}\neq y_j$.

א. נניח בשלילה שיש פתרון אופטימלי לבעיה כאשר
מוותרים כאשר יש j עבור $\epsilon_j<0$. יהיו $\epsilon_1,\dots,\epsilon_m$
ונגדיר $\epsilon'_1,\dots,\epsilon'_m$:

$\epsilon'_j=0$, $\epsilon'_i=\epsilon_i$: $\forall i\neq j$

מבחינת אילוצים - כל אילוץ שלא כולל את ϵ_j
נשאר אותו דבר, ולגבי האילוץ האחר:

$$y_j(w^Tx_j+b)\geq 1-\epsilon_j \implies y_j(w^Tx_j+b)\geq 1-\epsilon'_j$$

כי $1-\epsilon'_j=0$ אבל $1-\epsilon_j>1$ כי $\epsilon_j<0$.
ומבחינת הערך האופטימלי ברור ש- $\frac{C}{2}\sum_{i=1}^m \epsilon_i'^2 < \frac{C}{2}\sum_{i=1}^m \epsilon_i^2$
כי לכל $i\neq j$ $\epsilon_i^2=\epsilon_i'^2$ ו- $\epsilon_j'^2=0<\epsilon_j^2$.
כלומר מצאנו פתרון יותר טוב מ- $w,b,\epsilon_1,\dots,\epsilon_m$
בסתירה.

ב. הלגרנז'יאן הוא:

$$L(w,b,\epsilon,\lambda)=\frac{1}{2}w^Tw+\frac{C}{2}\sum_{i=1}^m\epsilon_i^2+\sum_{i=1}^m\lambda_i(1-\epsilon_i-y_i(w^Tx_i+b))$$

ג. נגזור ונשווה ל-0:

$$0=\nabla_w L(w,b,\epsilon,\lambda)=w+\sum_{i=1}^m-\lambda_i y_i x_i \implies w=\sum_{i=1}^m\lambda_i y_i x_i$$

$$0=\nabla_b L(w,b,\epsilon,\lambda)=\sum_{i=1}^m-\lambda_i y_i \implies \sum_{i=1}^m\lambda_i y_i=0$$

$$0=\nabla_{\epsilon_i} L(w,b,\epsilon,\lambda)=C\epsilon_i-\lambda_i \implies \epsilon_i=\frac{\lambda_i}{C}$$

נציב חזרה ונקבל:

$$L(w,b,\epsilon,\lambda)=\frac{1}{2}\left(\sum_{i=1}^m\lambda_i x_i y_i\right)^2+\frac{C}{2}\sum_{i=1}^m\frac{\lambda_i^2}{C^2}+\sum_{i=1}^m\lambda_i-\frac{\lambda_i^2}{C}=$$

$$=\frac{1}{2}\left(\sum_{i=1}^m\lambda_i x_i y_i\right)^2-\frac{1}{2C}\sum_{i=1}^m\lambda_i^2+\sum_{i=1}^m\lambda_i$$

ד. הבעיה הדואלית:

$$\max_{\lambda}\frac{1}{2}\left(\sum_{i=1}^m\lambda_i x_i y_i\right)^2-\frac{1}{2C}\sum_{i=1}^m\lambda_i^2+\sum_{i=1}^m\lambda_i$$

תחת האילוצים:

$$\sum_{i=1}^m\lambda_i y_i=0$$