~~שאלה 1~~

~~(a) יהיו $x,y\in\mathbb{R}^n$ ונוכיח שמתקיים: $g(\lambda x+(1-\lambda)y)\le\lambda g(x)+(1-\lambda)g(y)$~~

שאלה 1

(a) יהיו $x,y\in\mathbb{R}^n$, $\lambda\in\mathbb{R}$ ונוכיח שמתקיים:

$$g(\lambda x+(1-\lambda)y)\le\lambda g(x)+(1-\lambda)g(y).$$

$$g(\lambda x+(1-\lambda)y)=f(A(\lambda x+(1-\lambda)y)+b)\underset{b=\lambda b+(1-\lambda)b}{=}f(\lambda(Ax+b)+(1-\lambda)(Ay+b))\underset{f\text{ קמורה}}{\le}\lambda f(Ax+b)+(1-\lambda)f(Ay+b)=$$

$$\underset{\text{מהגדרת } g}{=}\lambda g(x)+(1-\lambda)g(y)$$

ולכן g קמורה, כנדרש!

$\square$

(b) נוכיח באופן דומה, לפי הגדרה: יהיו $x,y\in\mathbb{R}^n$, $\lambda\in\mathbb{R}$ ומתקיים:

$$g(\lambda x+(1-\lambda)y)=\max_i f_i(\lambda x+(1-\lambda)y)\underset{f_i\text{ קמורה}}{=}\max_i \lambda f_i(x)+\max_i (1-\lambda)f_i(y)=\underbrace{\lambda\cdot g(x)}+(1-\lambda)g(y)$$

ולכן עפ"י ההגדרה g קמורה.

$\square$

(c) נראה שמתקיים $\text{llog}''>0$ לכל $z\in\mathbb{R}$ אז llog קמורה.

$$\text{llog}'(z)=\frac{1-e^{-z}}{1+e^{-z}}$$

$$\text{llog}''(z)=\frac{e^{-z}(1+e^{-z})+e^{-z}(1-e^{-z})}{(1+e^{-z})^2}=\frac{e^{-z}(1+\cancel{e^{-z}}+1-\cancel{e^{-z}})}{(1+e^{-z})^2}=\frac{2e^{-z}}{(1+e^{-z})^2}>0$$

ולכן $\text{llog}''(z)>0$ לכל $z\in\mathbb{R}$, ולכן הפונקציה קמורה! ☐

~~(d) נראה ש $f(w)=\log_2(\dots)$ קמורה (מעל ... משתנים).~~

~~$f(w)=\sum_{i=1}^{n}\log_2(1+e^{-y_i w x_i})$~~

~~(d) ... קמורה ... $f(w)$ ...~~

~~... (1)~~

~~לכן ... $f(w)$ קמורה, כנדרש! ☐~~

(d) $f(w)=\log(y\langle w,x\rangle)$ נשים לב שזו הרכבה של 2 פונקציות קמורות ב-w. llog קמורה

כמו שראינו, ו: $g(w)=y\langle w,x\rangle$ אשר לינארית:

$$g(\lambda u+(1-\lambda)h)=y\langle w,\lambda u+(1-\lambda)h\rangle=y\langle w,\lambda u\rangle+y\langle w,(1-\lambda)h\rangle=\lambda g(u)+(1-\lambda)g(h)$$

כנדרש, ולכן

ולכן $g(w)$ קמורה, ולכן: $f=\text{llog}\circ g$ ולכן $f(w)$ קמורה. ☐

שאלה 2

ראשית, נוכיח כי 0-1 loss חסום מלמעלה ע"י hinge.

תזכורת: $\ell_{hinge}(y_i \cdot \langle w, x_i\rangle) = \max\{0, 1 - y_i \cdot \langle w, x_i\rangle\}$

$$\ell_{0-1} = \begin{cases} 0 & sign(y_i) = sign(\langle w, x_i\rangle) \\ 1 & else \end{cases}$$

אם הסימנים שונים, 0-1 loss יחזיר 1, אך: $1 - y_i\langle w, x_i\rangle > 1$ (כי <0)

ולכן hinge loss יחזיר מספר גדול מ-1.

כאשר הסימנים שווים, 0-1 loss יחזיר 0 ו- hinge loss לא יכול אף פעם להיות קטן מ-0.

לכן, בכל מקרה hinge loss חוסם מלמעלה את 0-1 loss.

כעת, יהי w^* המסווג, ויהי $c>0$, מתקיים:

$$\ell_{hinge}(y_i \cdot c \cdot \langle w^*, x_i\rangle) = \max\{0, 1 - c \cdot y_i \cdot \langle w^*, x_i\rangle\} \xrightarrow{c\to\infty} 0$$

כיוון ש- $\langle x_i, w^*\rangle y_i > 0$ לכל i, וזה מספר סופי על כל הדוגמאות, ולכן ככל ש-c שואף לאינסוף, נקבל ביטוי שלילי של מכפלה גדולה מאוד שלילי, ולכן יחזיר 0 ה-hinge loss.

ולכן מתקיים: $0 \le \ell_{0-1} \le \ell_{hinge} \xrightarrow{c\to\infty} 0$

ובסוף הכול, ב-$c\to\infty$ לא יתכן אחרת מ-0 ל-0-1 loss על w^*, ולכן זה נדרש!

$\square$

שאלה 3

(a) יהי $\lambda \in (0,1)$ ויהי $z \in K$, x מקמירות נקבל:

$$k = (1-\lambda)x + \lambda z = x - \lambda(x-z)$$

$$\|x-y\|_2^2 \le \|x-\lambda(x-z)-y\|_2^2 = \|(x-y)-\lambda(x-z)\|_2^2 \le \|x-y\|_2^2 - 2\lambda\langle x-y, x-z\rangle + \lambda^2\|x-z\|_2^2$$

משולש:

לאחר העברת אגפים:

$$\langle x-y, x-z\rangle \le \frac{\lambda}{2}\|x-z\|_2^2 .$$

זה נכון לכל $\lambda \in (0,1)$, ולכן אם נשאיף את λ ל-0 נקבל שמתקיים

$0 \ge \langle x-y, x-z\rangle$. כעת, נשים לב כי

לכן, נשתמש בזה ונקבל:

$$\|y-z\|_2^2 - \|x-z\|_2^2 = \|y-z+x-x\|_2^2 - \|x-z\|_2^2 = \|(x-z)-(x-y)\|_2^2 - \|x-z\|_2^2 =$$

$$= \cancel{\|x-z\|_2^2} - 2\langle x-y, x-z\rangle + \|x-y\|_2^2 - \cancel{\|x-z\|_2^2} \ge 0$$

$-2\langle x-y, x-z\rangle \ge 0$ כי הראנו שזה מתקיים מעל

ולכן קיבלנו: $\|y-z\|_2^2 \ge \|x-z\|_2^2$, כנדרש!

□

(b) נוכיח כ-(a) באופן דומה

$$\|x_{t+1}-x^*\|_2^2 = \|\Pi_K(y_{t+1})-x^*\|_2^2 \underset{(CS)}{\le} \|y_{t+1}-x^*\|_2^2 = \|x_t-\eta\nabla F(x_t)-x^*\|_2^2 =$$

$$= \|x_t-x^*\|_2^2 - 2\eta\nabla F(x_t)(x_t-x^*)+\eta^2\|\nabla F(x_t)\|_2^2$$

כמו בסעיף !!

$$\nabla F(x_t)\cdot(x_t-x^*) = \frac{\|x_t-x^*\|_2^2-\|x_{t+1}-x^*\|_2^2}{2\eta}+\frac{\eta}{2}\|\nabla F(x_t)\|_2^2$$

$$F(\bar{x})-F(x^*) \le \frac{1}{T}\sum_{t=1}^{T}\nabla F(x_t)\cdot(x_t-x^*) = \sum_{t=1}^{T}\left(\frac{1}{2\eta}\cdot\left(\|x_t-x^*\|_2^2-\|x_{t+1}-x^*\|_2^2\right)+\frac{\eta}{2}\|\nabla F(x_t)\|_2^2\right)=$$

כמו בסעיף (קמור)
ואז טלסקופי

$$= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\eta}{2}\|\nabla F(x_t)\|_2^2\right)+\frac{\|x_1-x^*\|_2^2}{2\eta T} \le \frac{\eta}{2}G^2+\frac{\|w^*\|_2^2}{2\eta T} \le \frac{\eta}{2}G^2+\frac{B^2}{2T\eta}=\epsilon$$

מנתונים $x_1=0$ כמו בסעיף

ולכן עבור $\eta=\frac{\epsilon}{G^2}$ ועבור $T=\frac{B^2G^2}{\epsilon^2}$ נקבל את הדרוש!!

$\square$

שאלה 4

GD מעדכון:

$x_{t+1} = x_t - \eta \nabla f(x_t) \Rightarrow x_{t+1} - x_t = -\eta \nabla f(x_t)$ (*)

לכן נציב בחסם העליון הנובע מחלקות את (*) ונקבל:

$$f(x_{t+1}) \le f(x_t) + \langle \nabla f(x_t)^T, -\eta \nabla f(x_t) \rangle + \frac{\beta \eta^2}{2} \|\nabla f(x_t)\|^2$$

$$\Downarrow$$

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

$$\Downarrow \quad \eta - \frac{\beta^2 \eta^2}{2} > 0 \Leftarrow \eta < \frac{2}{\beta}$$

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

$$\sum_{t=1}^{k} \|\nabla f(x_{t+1})\|^2 \le \sum_{t=1}^{k} \frac{f(x_t) - f(x_{t+1})}{\eta - \frac{\beta^2 \eta^2}{2}} = \sum_{t=1}^{k} \frac{f(x_1) - f(x_{k+1})}{\eta - \frac{\beta^2 \eta^2}{2}} \le \frac{f(x_1)}{\eta - \frac{\beta^2 \eta^2}{2}} < \infty$$

סכום טלסקופי

אם נגדיר $k \to \infty$:

$f(x) \ge 0$ נתון

טור מונוטוני חסום מתכנס לגבול סופי, וכיוון שכל איברי הטור

ולכן בהכרח, מתקיים:

$$\lim_{t \to \infty} \|\nabla f(x_t)\| = 0$$

$\square$

שאלה [illegible]

[illegible]

1) a. [illegible] σ_f [illegible] 1.

b. [illegible]

c. [illegible]

d. [illegible]:

[illegible]

[illegible]

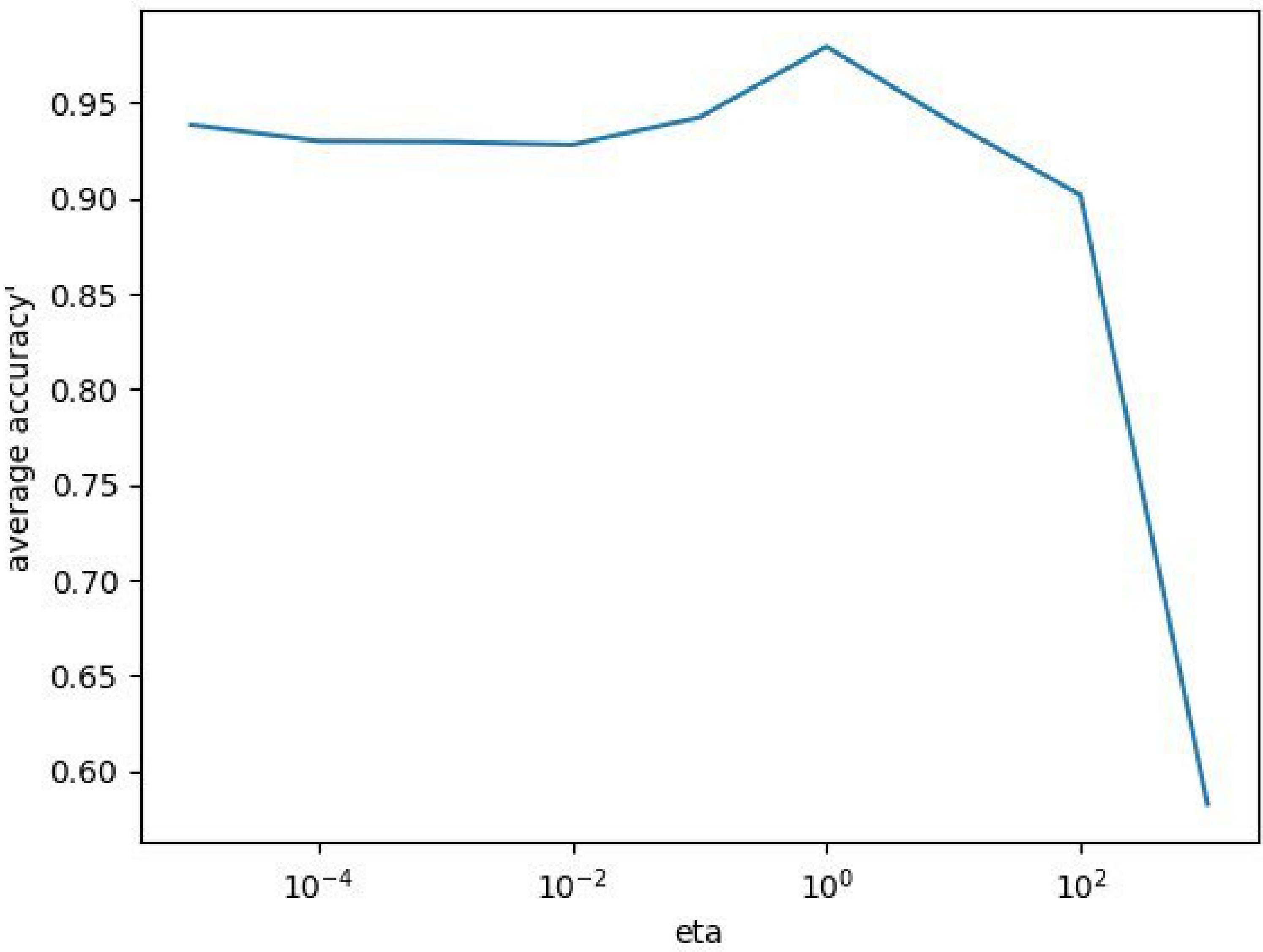
2) a. [illegible] σ_f [illegible]

b. [illegible]:

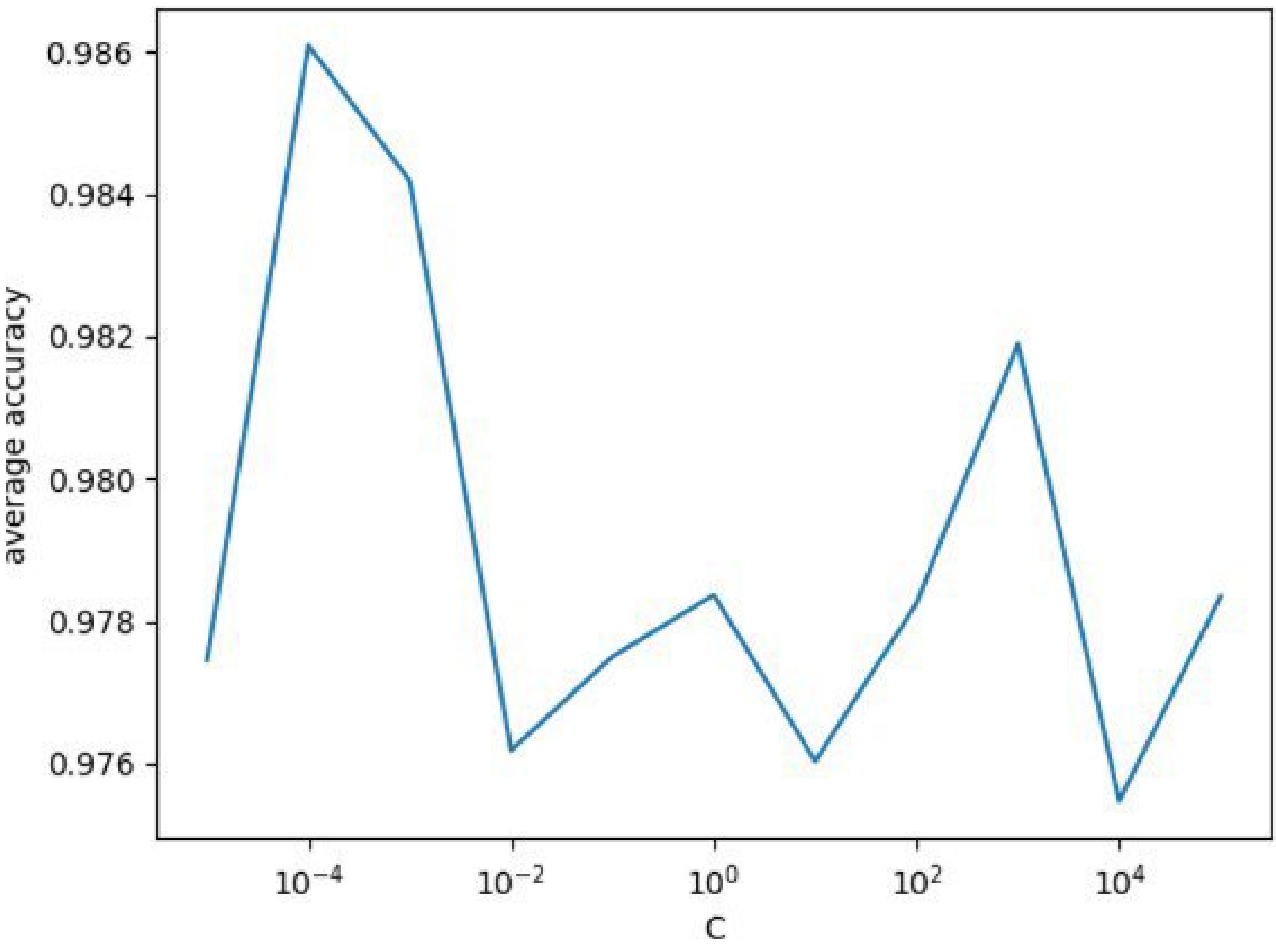
[illegible]

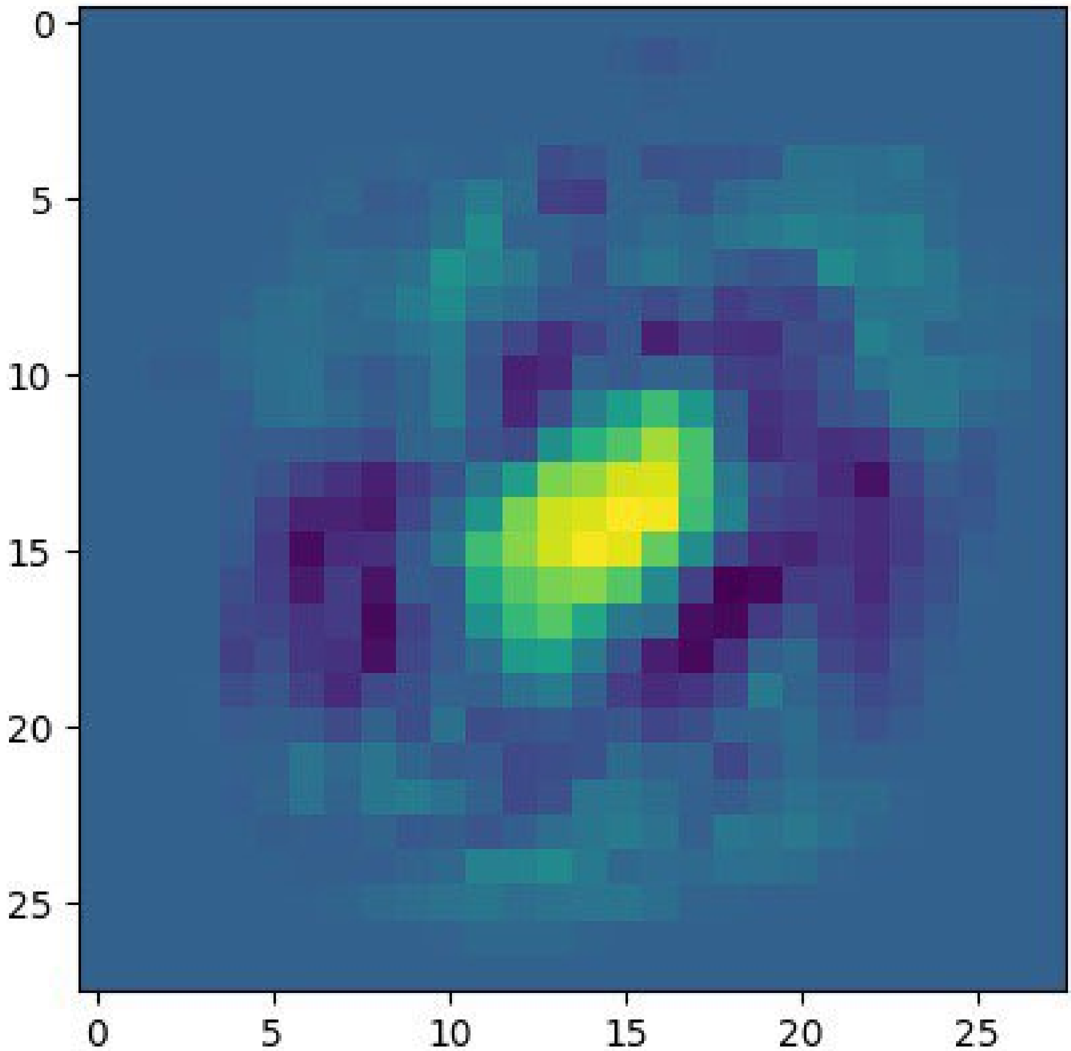
[illegible]

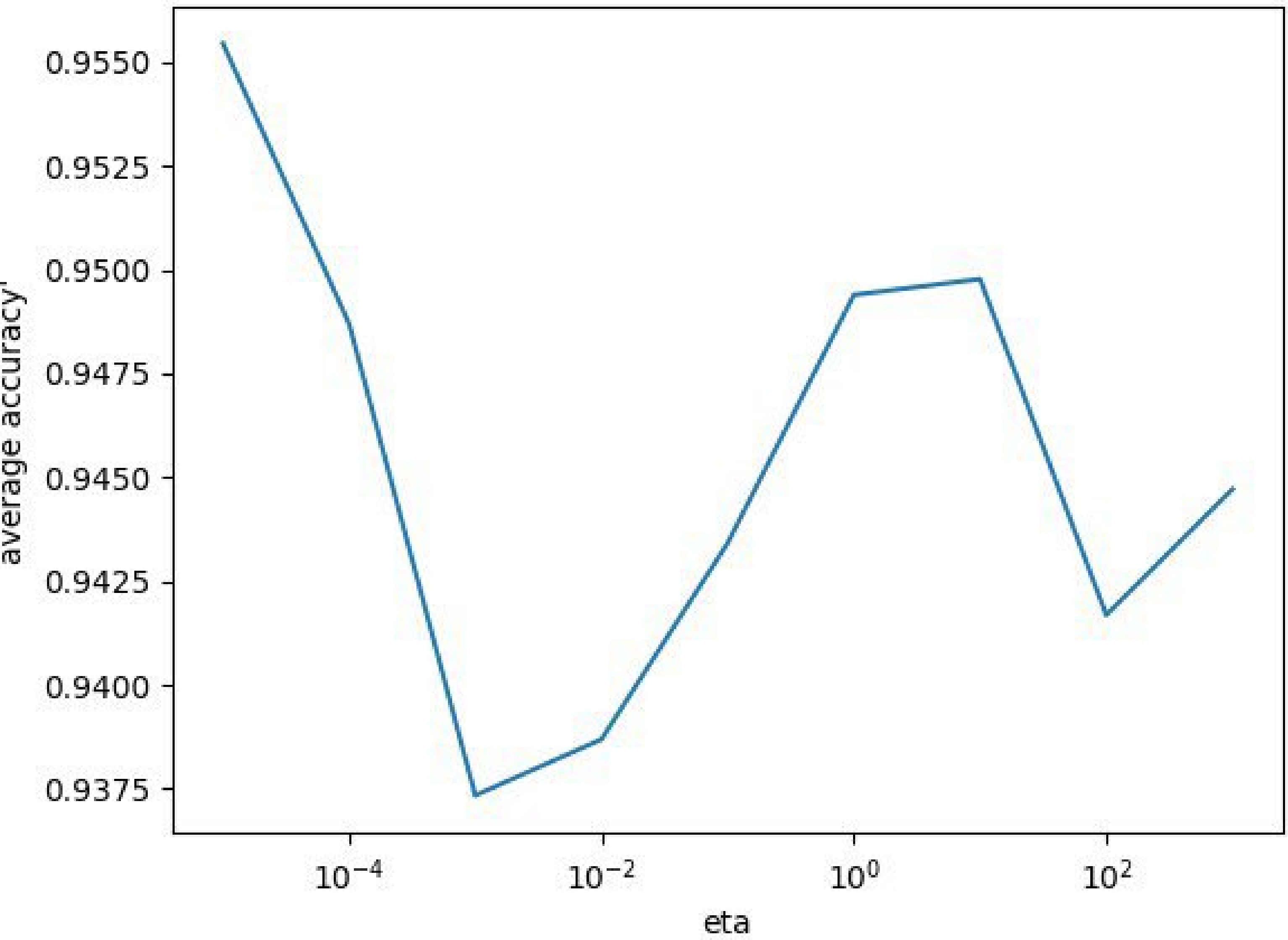
c. [illegible]



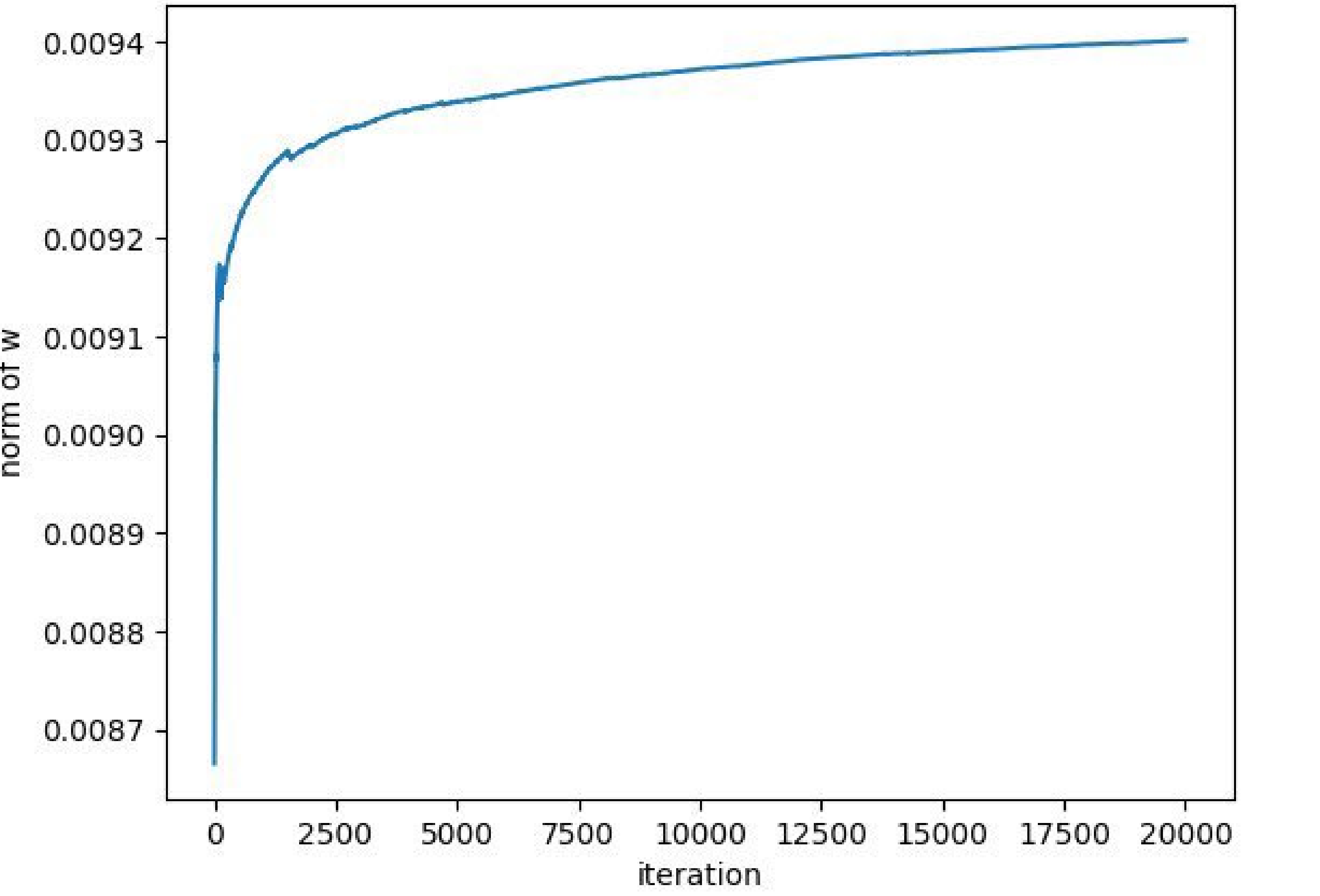
average accuracy'
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
10^{-4}
10^{-2}
10^{0}
10^{2}
eta

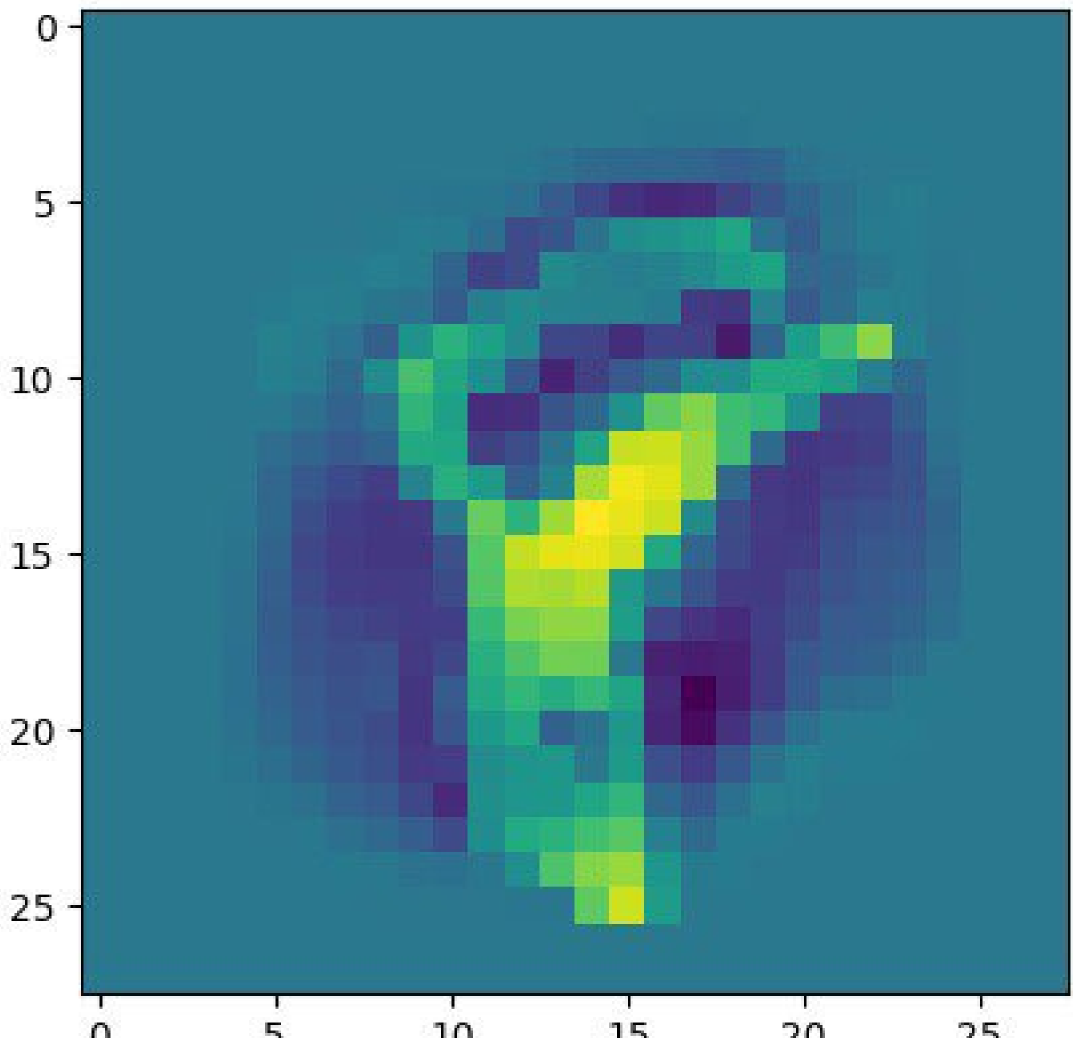

0.986
0.984
0.982
0.980
0.978
0.976
average accuracy
10^{-4}
10^{-2}
10^{0}
10^{2}
10^{4}
C


0
5
10
15
20
25
0
5
10
15
20
25



0.9550
0.9525
0.9500
0.9475
0.9450
0.9425
0.9400
0.9375
average accuracy
10^{-4}
10^{-2}
10^{0}
10^{2}
eta


0.0094
0.0093
0.0092
0.0091
0.0090
0.0089
0.0088
0.0087
norm of w
0
2500
5000
7500
10000
12500
15000
17500
20000
iteration


0
5
10
15
20
25
0
5
10
15
20
25