תרגיל 5

שם: תמר שבן
ת.ז: 315315523

1) יהי psd kernel k, נגדיר פונקציית קרנל $\tilde{k}$ ע"י:

$$\tilde{k}(x,x') = \frac{k(x,x')}{\sqrt{k(x,x)\cdot k(x',x')}}$$

ראשית נשים לב כי זוהי אכן פונקציית קרנל שכן k פונקציית קרנל.
כעת נוכיח כי זוהי פונק קרנל PSD.

מהנתון, ממרסר, נובע ש k PSD. קיימת $\psi : X \to F$ כך ש

$$k(x,x') = \langle \psi(x), \psi(x') \rangle$$

כעת נקבל

$$\tilde{k}(x,x') = \frac{k(x,x')}{\sqrt{k(x,x)\cdot k(x',x')}}$$

$$= \frac{\langle \psi(x), \psi(x')\rangle}{\sqrt{\|\psi(x)\|^2}\cdot\sqrt{\|\psi(x')\|^2}}$$

כאשר $\hat{\psi}(x) = \frac{\psi(x)}{\|\psi(x)\|}$

$$= \frac{\langle \psi(x), \psi(x')\rangle}{\|\psi(x)\|\cdot\|\psi(x')\|} = \frac{\psi(x)^t\,\psi(x')}{\|\psi(x)\|\cdot\|\psi(x')\|} = \hat{\psi}(x)^t\cdot\hat{\psi}(x')$$

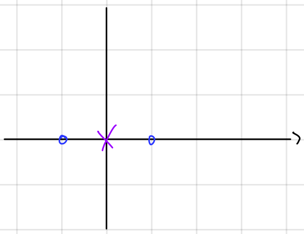
$$= \langle \hat{\psi}(x), \hat{\psi}(x') \rangle$$

ולכן מתנאי מרסר קיבלנו כי $\tilde{k}$ הינו קרנל PSD
ובנוסף לכל x נקבל

$$\tilde{k}(x,x) = \frac{k(x,x)}{\sqrt{k(x,x)\cdot k(x,x)}} = \frac{k(x,x)}{k(x,x)} = 1$$

2) $S = \{(x_i, y_i)\}_{i=1}^{m}$ $y_i \in \{\pm 1\}$

$$S = \{((-1,0),1),((0,0),-1),((1,0),1)\} \quad d=2$$



$$\psi(x_1,x_2) = (x_1, x_2, x_1^2+x_2^2) \quad F = \mathbb{R}^3$$

כעת כפי שראינו בתרגול, הדגימות בעלות הסימון 1 יקבלו ערך חיובי בקורדינטה השלישית ולכן ירחקו מהמישור $span\left(\begin{bmatrix}1\\0\\0\end{bmatrix},\begin{bmatrix}0\\1\\0\end{bmatrix}\right)$ בעוד הנקודה $(0,0)$ תישאר בראשית הצירים.
ולכן קיבלנו שהמישור $span\left(\begin{bmatrix}1\\0\\0.5\end{bmatrix},\begin{bmatrix}0\\1\\0.5\end{bmatrix}\right)$ מפריד בין הנקודה $\begin{bmatrix}0\\0\\0\end{bmatrix}$ לבין הנקודות $\begin{bmatrix}1\\0\\1\end{bmatrix}$ ו $\begin{bmatrix}0\\1\\1\end{bmatrix}$

3) א) יהי $K(x,y)$ פונק' קרנל ! K מוגדר הגרם שלה.
יהיו N נקודות. צ"ל קיימת $f:X\to\mathbb{R}^N$ כך ש $k(x_i,x_j)=f^T(x_i)f(x_j)$
בהינתן N דגימות $x_1 \dots x_N$ ופונקציית psd קרנל מהנתון מספר קיימת f עבורה

$$k(x_i,x_j)=\langle f(x_i),f(x_j)\rangle = f(x_i)^t f(x_j)$$

כעת נשים ל k_2 גם psd קרנל קיימת g כך ש $k_2(x_i,x_j)=g(x_i)^t g(x_j)$

כעת
$$k_1(x,y)\cdot k_2(x,y)=f(x)^t f(y)\cdot g(x)^t g(y)$$
$$=Tr\left[f(x)^t f(y)\, g(x)^t g(y)\right]$$
$$=Tr\left[g(y)\, f(x)^t f(y)\, g(x)^t\right]$$
$$=\sum_i\left[g(y)f(x)^t f(y) g(x)^t\right]_{ii}=\sum_i\sum_j g_i(x)f_j(x)f_j(y)g_i(y)$$

ב) נגדיר $h:X\to\mathbb{R}^{N^2}$, בהינתן X, נסמן את הקורדינטה ה- של $h(x)$ כתור ij כאשר i ו-j מציינים מקום בגודל N ! נ"ב באורך הוקטור בגודל N -וצר כניסה. כעת

$$[h(x)]_{ij}=g_i(x)f_j(x)$$

כעת נשים ש $k_x(x,y)=k_1(x,y)k_2(x,y)=\sum_i\sum_j g_i(x)f_j(x)f_j(y)g_i(y)$
$$=\sum_i\sum_j g_i(x)f_j(x)g_i(y)f_j(y)$$
$$=h(x)^t\cdot h(y)=\langle h(x),h(y)\rangle$$

כלומר מהנתון נובע k_x פונק' קרנל psd

$$\Sigma = Cov(X) = \mathbb{E}\left((X-\mathbb{E}(X))(X-\mathbb{E}(X))^t\right) = \mathbb{E}(XX^t)$$

4) $X:\Omega \to \mathbb{R}^d$ נ"ל כך ש $\mathbb{E}(X)=\vec{0}$

יהי $v\in\mathbb{R}^d$ כך ש $\|v\|=1$ נגדיר את הבא $\hat{X} = \langle v, X\rangle$

נשים לב כי

$$\mathbb{E}(\hat{X}) = \langle v, \mathbb{E}(X)\rangle = \langle v, \vec{0}\rangle = 0$$

כעת

$$Var(\hat{X}) = \mathbb{E}\left((\hat{X}-\mathbb{E}(\hat{X}))^2\right) = \mathbb{E}(\hat{X}\cdot\hat{X}) = \mathbb{E}(\langle v,X\rangle\langle v,X\rangle)$$

$$= \mathbb{E}\left((v^t X)\cdot(X^t v)\right)$$

$$= v^t\,\mathbb{E}(XX^t)\,v$$

$$= v^t \Sigma v$$

$$\overset{(1)}{\leq} u_1^t \Sigma u_1$$

(1) שכן u_1 הוא הו"ע עבור הערך הגבוהה ביותר של Σ לכן PCA

1.3

1) יהיו $u,v\in C$ ו- $\alpha\in[0,1]$ כעת

$$g(\alpha v+(1-\alpha)u)=\sum_{i=1}^{m}\gamma_i f_i(\alpha v+(1-\alpha)u)\overset{f_i\text{ קמורה}}{\le}\sum_{i=1}^{m}\gamma_i\left(\alpha f_i(v)+(1-\alpha)f_i(u)\right)$$
$$=\alpha\sum_{i=1}^{m}\gamma_i f_i(v)+\gamma_i(1-\alpha)f_i(u)$$
$$=\alpha\sum_{i=1}^{n}\gamma_i f_i(v)+(1-\alpha)\sum_{i=1}^{m}\gamma_i f_i(u)$$
$$=\alpha g(v)+(1-\alpha)g(u)$$

2) קודם נראה: נגדיר $f=-x$ $g=x^2$ ראשית נשים לב כי שתיהן קמורות
אכן $f=-x$ לינארית, ו- $g'=2x$, $g''=2>0$ אך $f\circ g=-x^2$ המקיים $(f\circ g)'=-2x$
$(f\circ g)''=-2<0$
כלומר $f\circ g$ קעורה ##

3) $\Leftarrow$ נניח כי f קמורה
יהיו $(u,s),(v,t)\in epi$ וכן $\alpha\in[0,1]$ נראה $\alpha(u,t)+(1-\alpha)(v,s)\in epi(f)$

$$\alpha(u,t)+(1-\alpha)(v,s)=(\alpha u,\alpha t)+((1-\alpha)v,(1-\alpha)s)$$
$$=(\alpha u+(1-\alpha)v,\ \alpha t+(1-\alpha)s)$$

כעת נשים לב ש $(u,t)\in epi(f)$, $(v,s)\in$ נקבל $f(u)\le t$, $f(v)\le s$ כי

$$f(\alpha u+(1-\alpha)v)\overset{\text{קמירות}}{\le}\alpha f(u)+(1-\alpha)f(v)\le\alpha t+(1-\alpha)s$$

כלומר $\alpha(u,t)+(1-\alpha)(v,s)=(\alpha u+(1-\alpha)v,\ \alpha t+(1-\alpha)s)\in epi(f)$

כיוון שני: נניח כי $epi(f)$ קמור. יהיו $u,v\in C$ ו- $\alpha\in[0,1]$ כעת נסמן $f(u)=t$ $f(v)=s$
כך שמתקיים $f(u)\le t$ ו- $f(v)\le s$ כלומר $(u,t)\in epi(f)$, (v,s) לכן
כלומר $(\alpha u+(1-\alpha)v,\ \alpha t+(1-\alpha)s)\in epi(f)$ מהקמירות כלומר $f(\alpha u+(1-\alpha)v)\le\alpha t+(1-\alpha)s$

$$f(\alpha u+(1-\alpha)v)\le\alpha t+(1-\alpha)s=\alpha f(u)+(1-\alpha)f(v)$$

כלומר f קמורה

$f(u)=\sup_{i\in I}(f_i(u))$ $f:V\to\mathbb{R}$ $f_i:V\to\mathbb{R}$ (4

יהיו $u,v\in V$ ו- $\alpha\in[0,1]$ כאלה נתונים

מהגדרת את ה הגדרה
לקמירות f_i $\forall i\in I$

תכונות סכום לסופרמום
+ כפל ב 2 סופרמה
בגלל חיובי (עבור).

$$f(\alpha u+(1-\alpha)v)=\sup_{i\in I}f_i(\alpha u+(1-\alpha)v)\leq\sup_{i\in I}\alpha f_i(u)+(1-\alpha)f_i(v)\leq\alpha\sup_{i\in I}f_i(u)+(1-\alpha)\sup_{i\in I}f(v)$$

$$=\alpha f(u)+(1-\alpha)f(v)$$

כלומר f קמורה $\boxtimes$

1.4

5) יהיו $(w,b),(w',b')$ וגם $\alpha\in[0,1]$ כעת

$$f(\alpha w+(1-\alpha)w',\ \alpha b+(1-\alpha)b') = \max(0,\ 1-y(x^T(\alpha w+(1-\alpha)w')+\alpha b+(1-\alpha)b'))$$

$$= \max(0,\ 1-y(\alpha x^Tw+(1-\alpha)x^Tw')+\alpha b+(1-\alpha)b'))$$

$$= \max(0,\ 1-y\alpha x^Tw-(1-\alpha)yx^Tw'-y\alpha b-y(1-\alpha)b'))$$

$$= \max(0,\ 1-\alpha y(x^Tw+b)-(1-\alpha)y(x^Tw'+b'))$$

$$= \max(0,\ \alpha(1-y(x^Tw+b))+(1-\alpha)(1-y(x^Tw'+b')))$$

$$\le \alpha\max(0,\ 1-y(x^Tw+b))+(1-\alpha)\max(0,\ 1-y(x^Tw'+b'))$$

$$\le \alpha f(w,b)+(1-\alpha)f(w',b')$$

- כיוון ש ב מקסימום גדול יותר - דוגמא: נסתכל החצה שבין ערך 0 לבין תוצאה נכפל ואז השפצה ועבור הגורם השני נסתכל ואם הוא חיובי

כנדרש ☒

6) נגדיר $h(w,b)=1-yx^Tw-yb$ כעת $\frac{\partial h}{\partial w}=-yx^T$ $\quad \frac{\partial h}{\partial b}=-y$

ולכן

$$g(w,b)=\begin{cases}0, & f(w,b)=0\\ (-yx^T,-y) & f(w,b)\neq 0\end{cases}$$

7) $f:\mathbb{R}^d\to\mathbb{R}$ $\quad f(x)=\sum_{i=1}^{m} f_i(x)$ $\quad$ נ"ל $\sum_k g_k\in\partial\sum_k f_k(x)$

מהגדרת תת-גרדיאנט לכל g_k של f_k מתקיים לכל $u\in\mathbb{R}^d$ בנקודה x

$$f_k(u)\ge f_k(x)+\langle g_k,u-x\rangle$$

כעת נסכום לכל k

$$\sum_{i=1}^{m} f_i(u)\ge\sum_{i=1}^{m} f_i(x)+\langle g_i,u-x\rangle=\sum_{i=1}^{m} f_i(x)+\sum_{i=1}^{m}\langle g_i,u-x\rangle$$

כלומר מהגדרה זה תת-גרדיאנט בנקודה ולכן $\sum_k g_k\in\partial\sum_k f_k(x)$

8)

$$f(w,b)=\frac{1}{m}\sum_{i=1}^{m}\ell^{hinge}_{x_i,y_i}(w,b)+\frac{\lambda}{2}\|w\|^2$$

נשתמש בסעיף הקודם הוא נסמן $f_i(w,b)=\ell^{hinge}_{x_i,y_i}(w,b)$ כעת

$$g_i(w,b)=\begin{cases}0 & ,f_i(w,b)=0\\(-yx^T,-y) & f_i(w,b)\neq 0\end{cases}$$

ולפי סעיף 7 נקבל כי $\sum_{i=1}^{m}g_i(w,b)\in\partial\sum_{i=1}^{n}f_i(w,b)$

כעת $\frac{\partial\frac{\lambda}{2}\|w\|^2}{\partial b}=0$ $\frac{\partial\frac{\lambda}{2}\|w\|^2}{\partial w}=\lambda w$

בסוף נקבל

$$\frac{1}{m}\sum_{i=1}^{m}g_i(w,b)+\lambda(w,0)\in\partial f(w,b)$$

כנדרש $\blacksquare$

2.1.1

א) ניתן לראות שב-L_2 הגרף של המודל היה בצורה עיגול ועבור L_1 הצורה היא מעוין ואפשר לראות את ההשפעה של זוית הנורמה למסלול של הגרדיאנט דיסנט.

ב) את הנ״ל שהבנו שכן אנו יודעים ש-L הנקודה הממוצעת יותר שווה בכל אינדקס, בנוסף ב-L_1 אנחנו רואים שהמסלול הולך לכיוון הפינה של המעוין כאשר ב-L_2 שמתקרב יותר עבור האפס.

ג) אנו יכולים לראות שעבור L_1 הירידה במודל היא ירידה לינארית שמגיעה ל-0 לאחר כמות ב-L_2 אנו רואים שהירידה מתחילה בצורה חדה ואז מתמתנת ומתקרבת ל-0. ניתן גם לראות את השוני בהגעה למינימום עבור ערכי eta שונים.

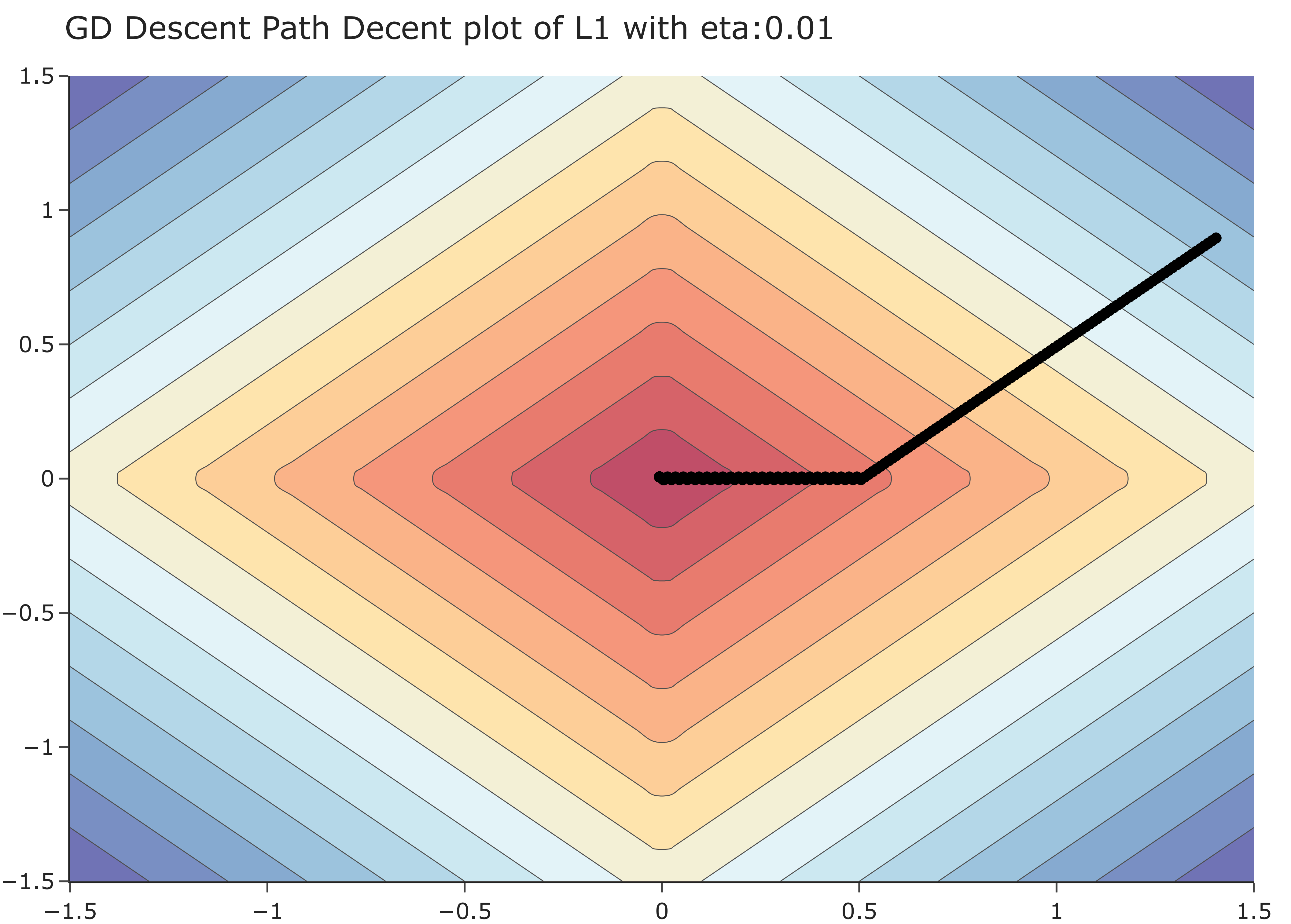
ד) הלוס המינימלי עבור L_1 הוא עבור eta-0.01 והוא: 0.008
הלוס המינימלי עבור L_2 הוא עבור eta-0.1 והוא: $1.4 \cdot e^{-9}$

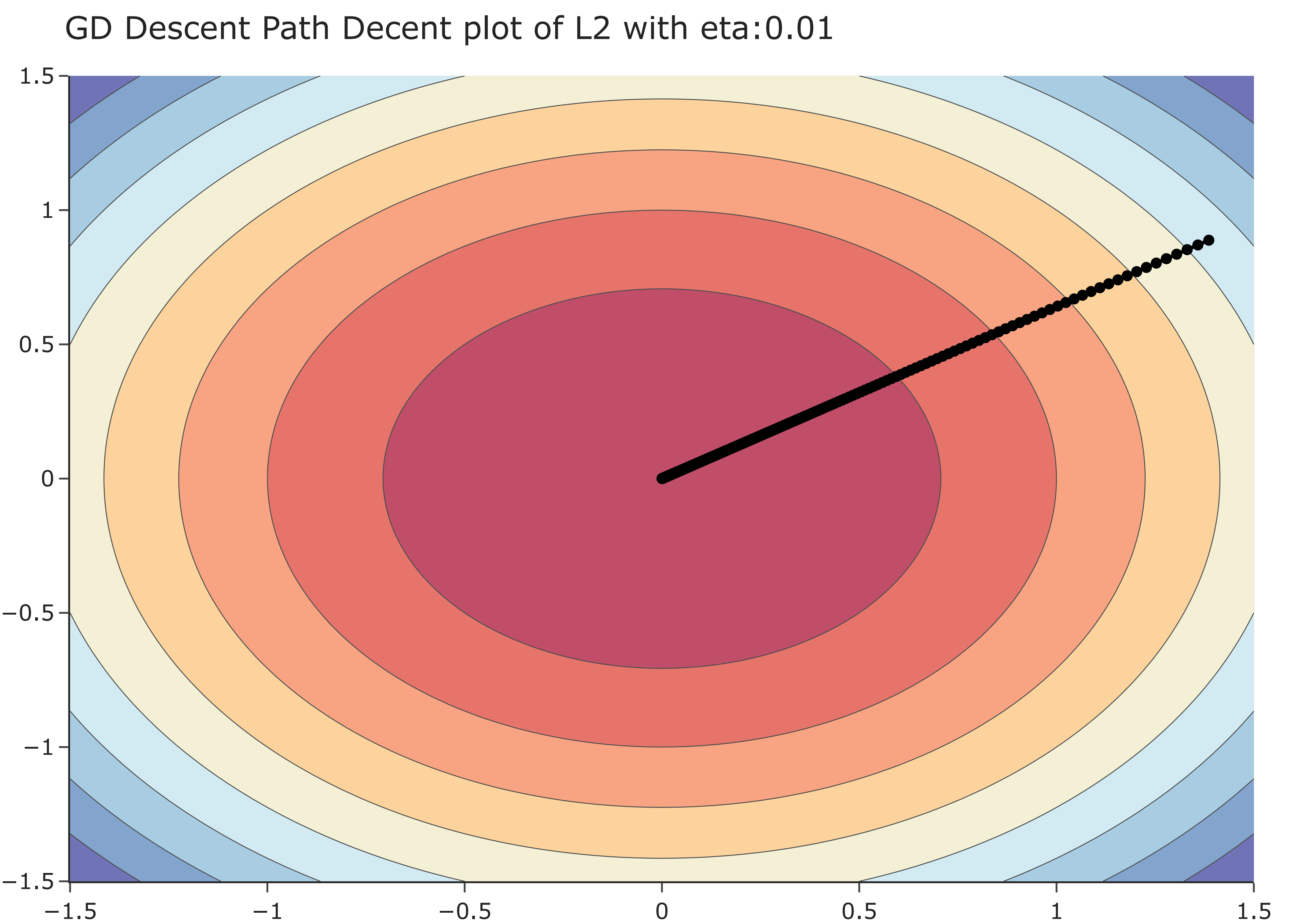
אנחנו יכולים לראות שעבור L_1 הערכים הקטנים הובילו למינימום הנמוך אבל עבור L_2 דווקא הערך הגבוה יותר עזר לנו להגיע למקום נמוך יותר בהרבה.

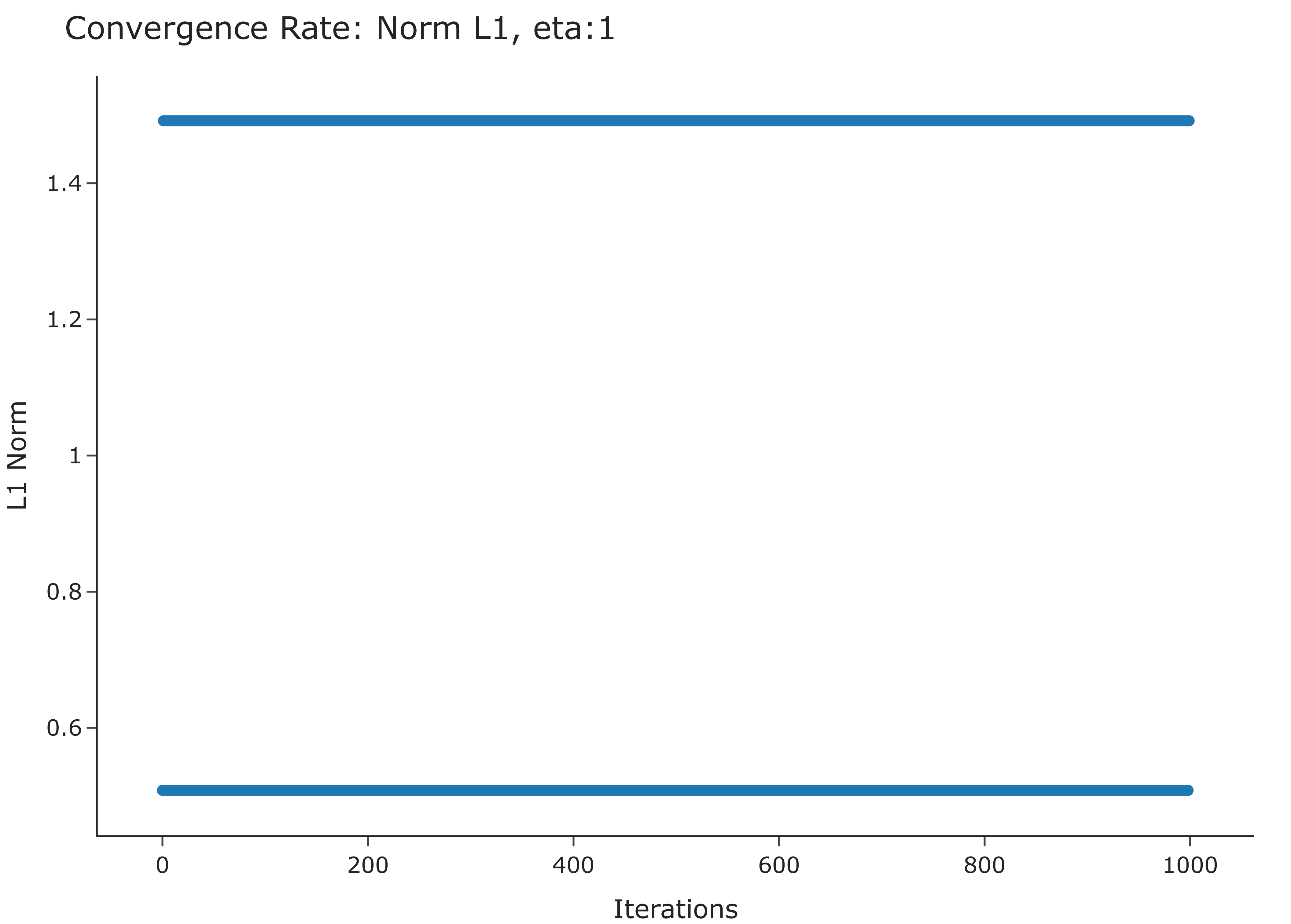
ט) ה-λ הכי טוב הוא 0.00013
והשגיאה של הטסט היא 0.36956

10) ערך ה-λ הטוב ביותר שקיבלתי עבור L_1 הוא 0.005 והלוס על הטסט היא 0.27173
יא) " " " " " " עבור L_2 " 0.05 " " " 0.260869

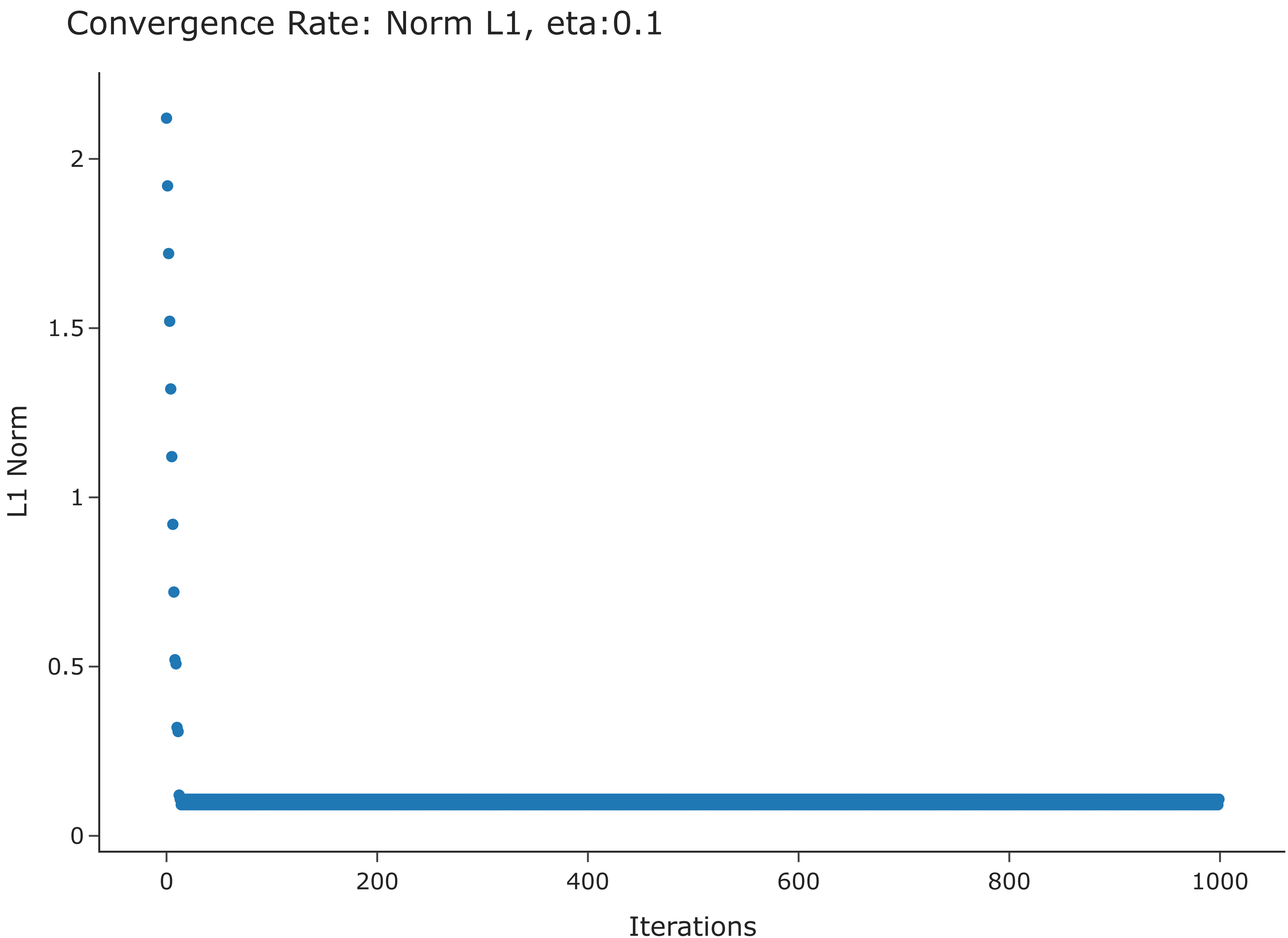
* הגרפים בדפים הבאים

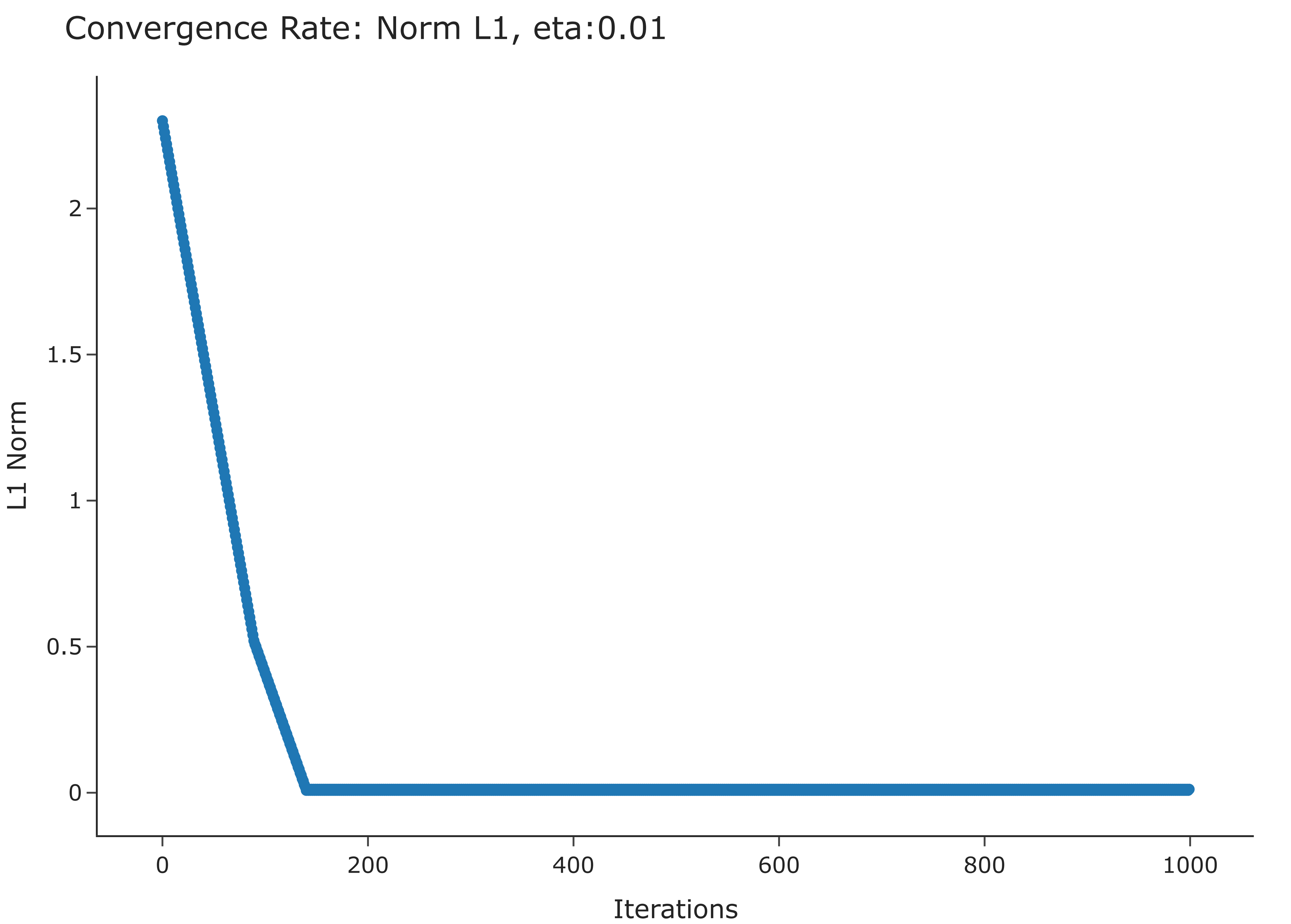

GD Descent Path Decent plot of L1 with eta:0.01
1.5
1
0.5
0
−0.5
−1
−1.5
−1.5
−1
−0.5
0
0.5
1
1.5

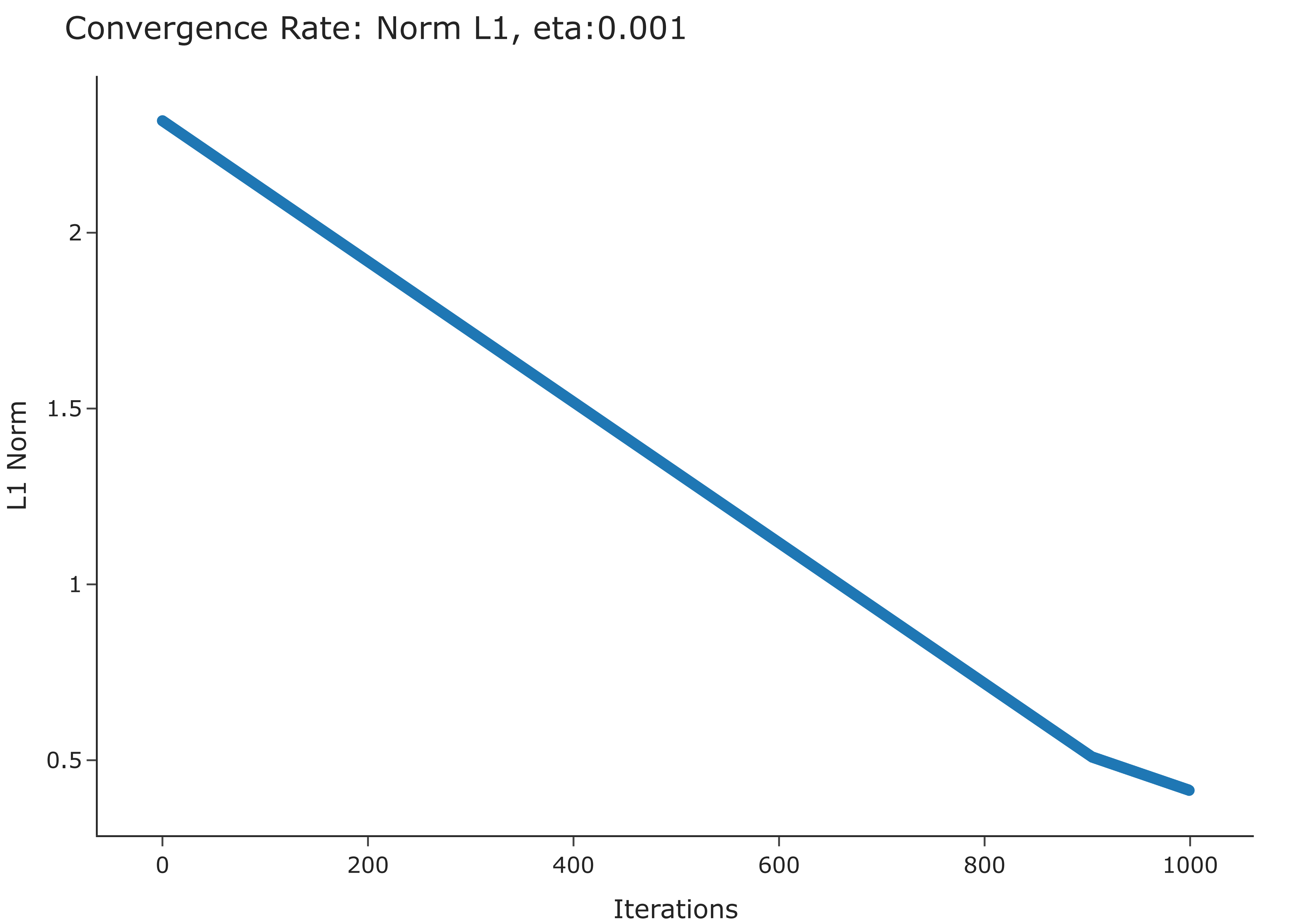

GD Descent Path Decent plot of L2 with eta:0.01
1.5
1
0.5
0
−0.5
−1
−1.5
−1.5
−1
−0.5
0
0.5
1
1.5



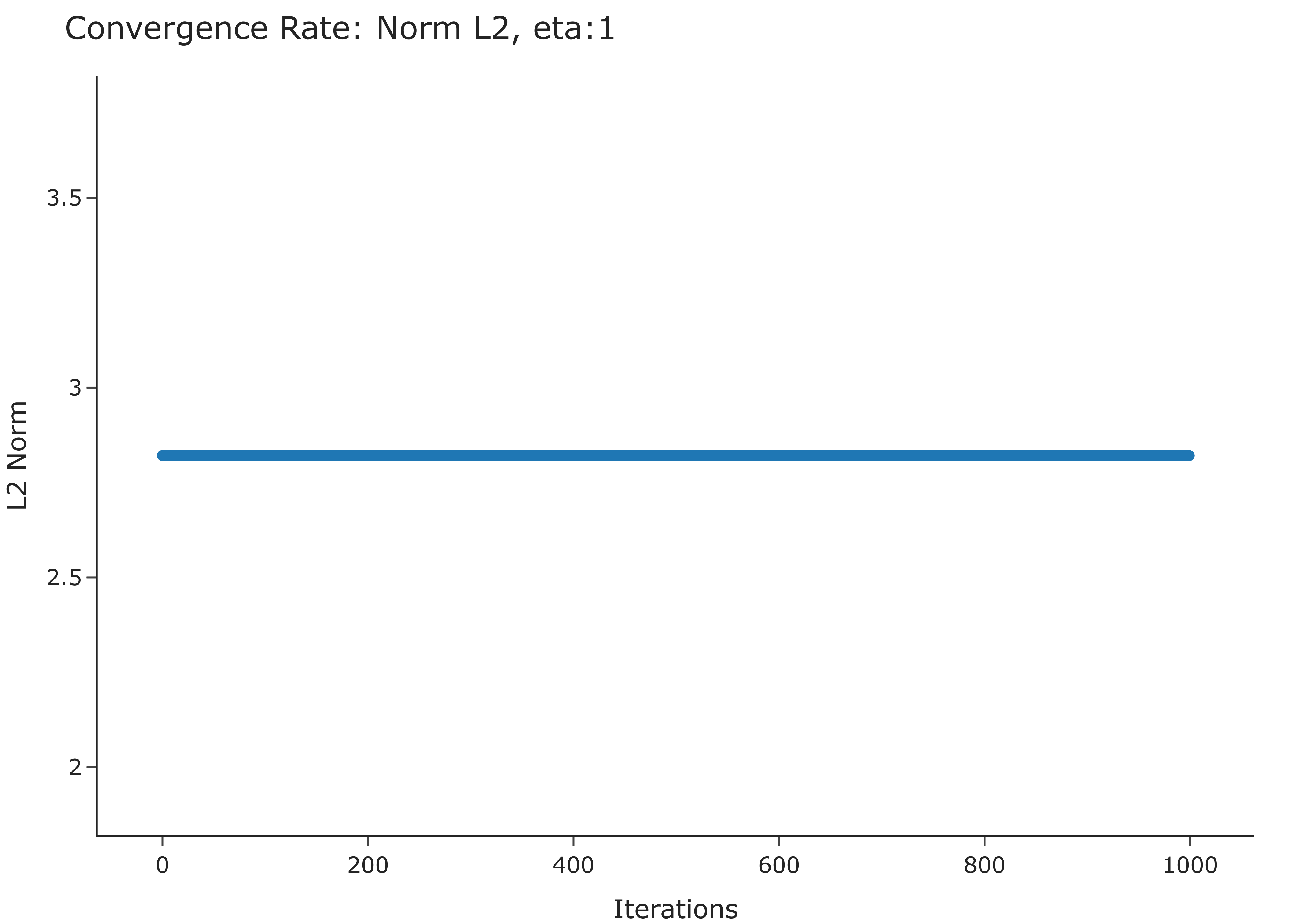
Convergence Rate: Norm L1, eta:1
L1 Norm
1.4
1.2
1
0.8
0.6
0
200
400
600
800
1000
Iterations

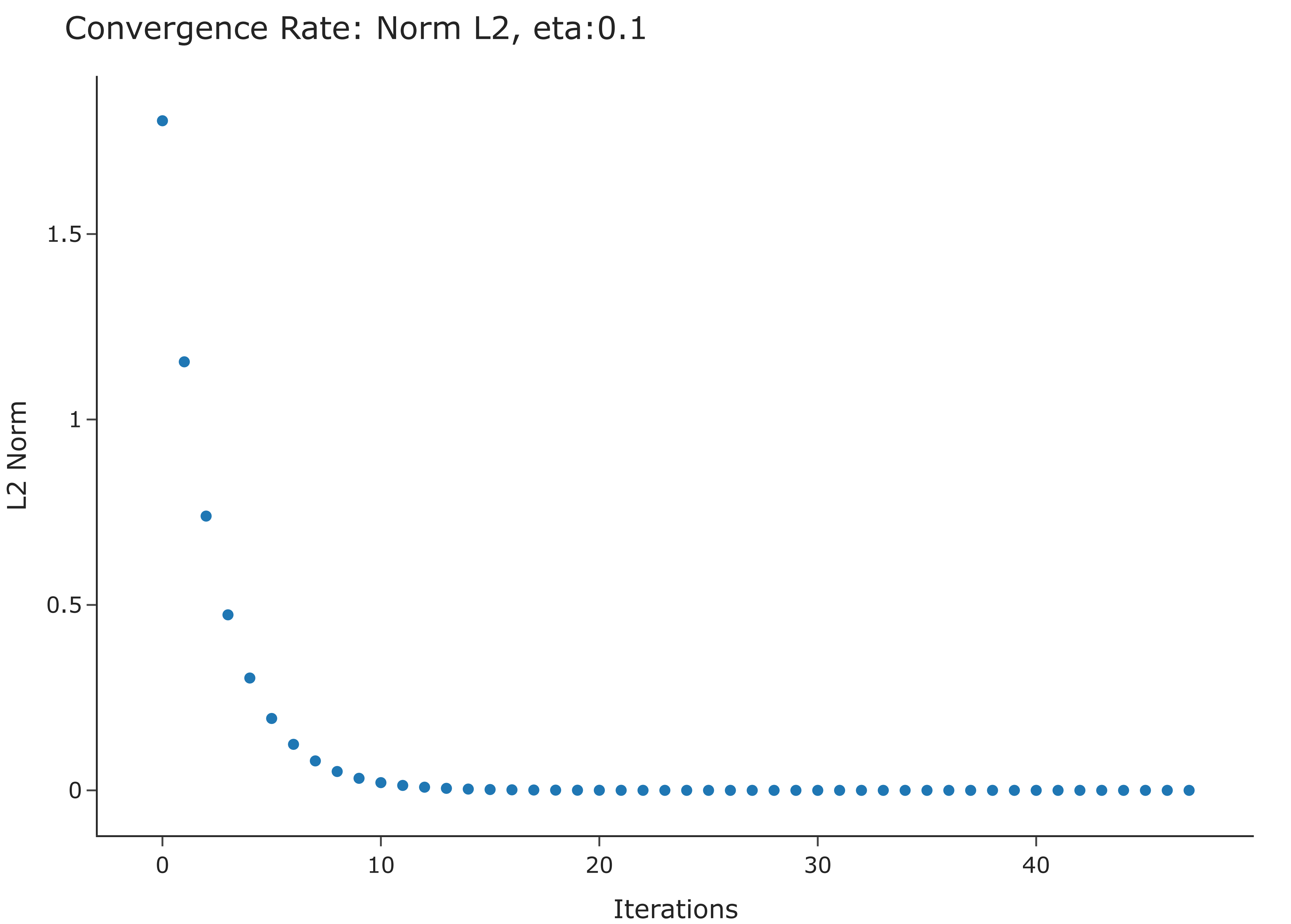

Convergence Rate: Norm L1, eta:0.1
L1 Norm
Iterations
0
0.5
1
1.5
2
0
200
400
600
800
1000

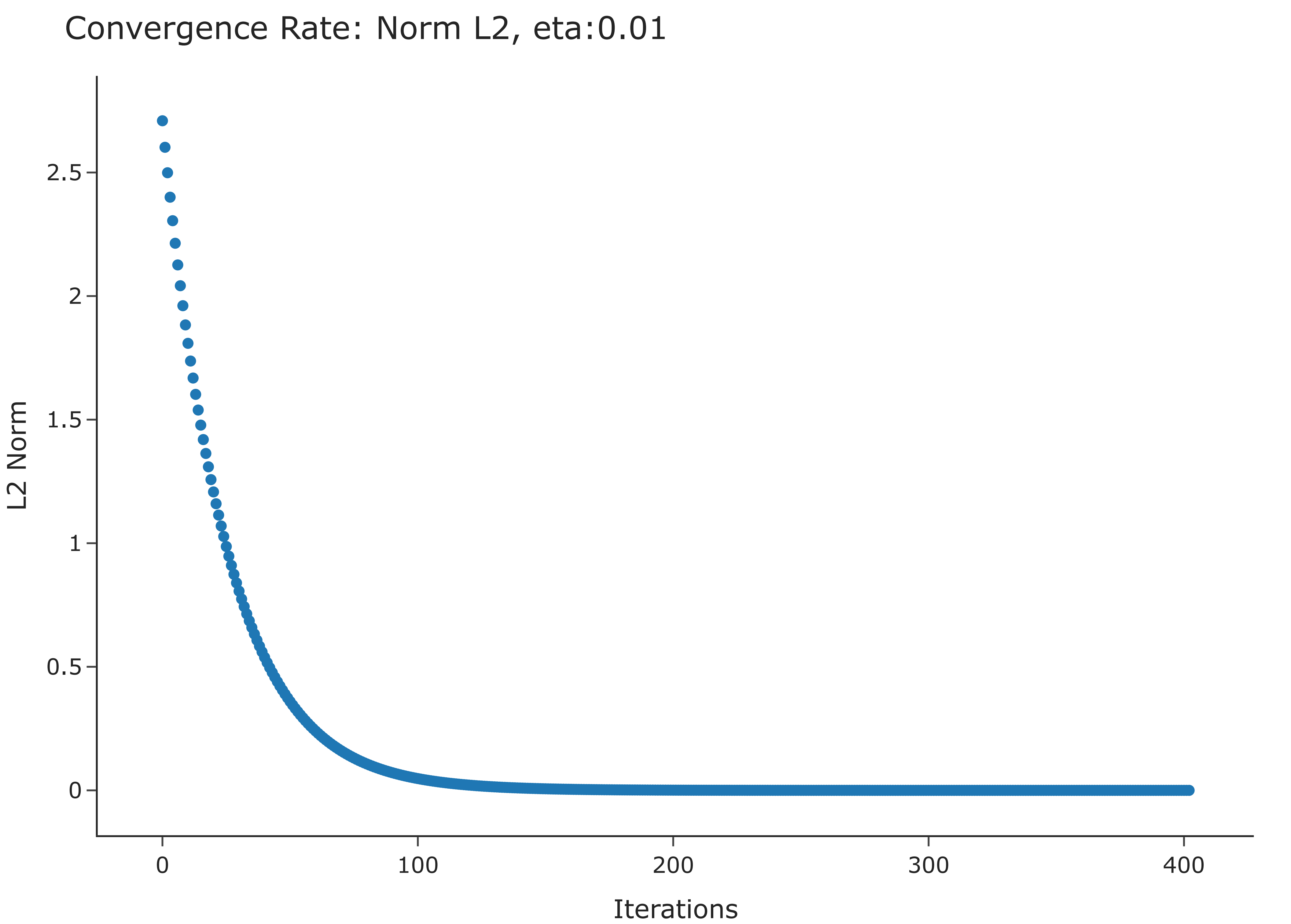

Convergence Rate: Norm L1, eta:0.01
L1 Norm
Iterations
0
0.5
1
1.5
2
0
200
400
600
800
1000



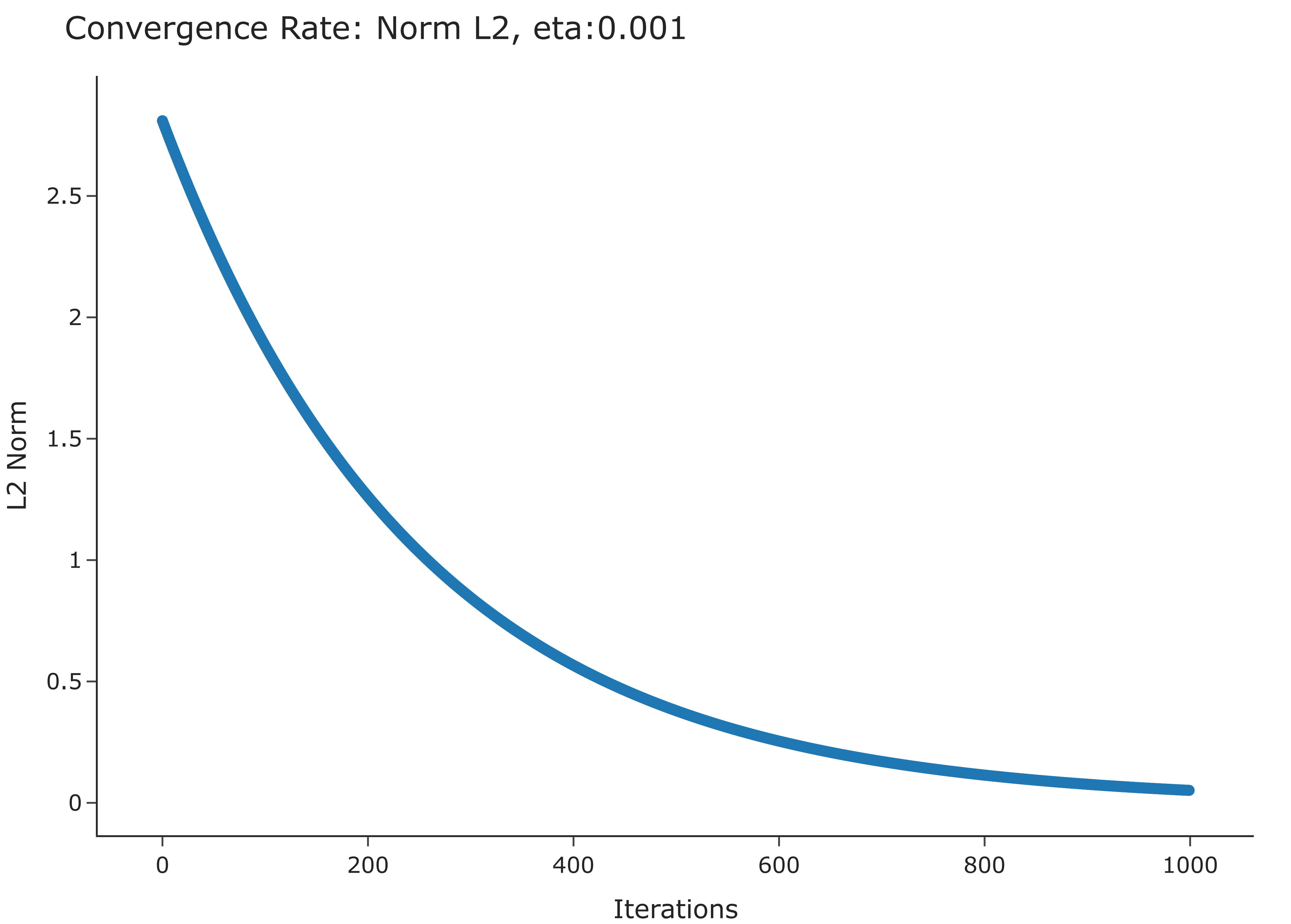
Convergence Rate: Norm L1, eta:0.001
L1 Norm
2
1.5
1
0.5
0
200
400
600
800
1000
Iterations

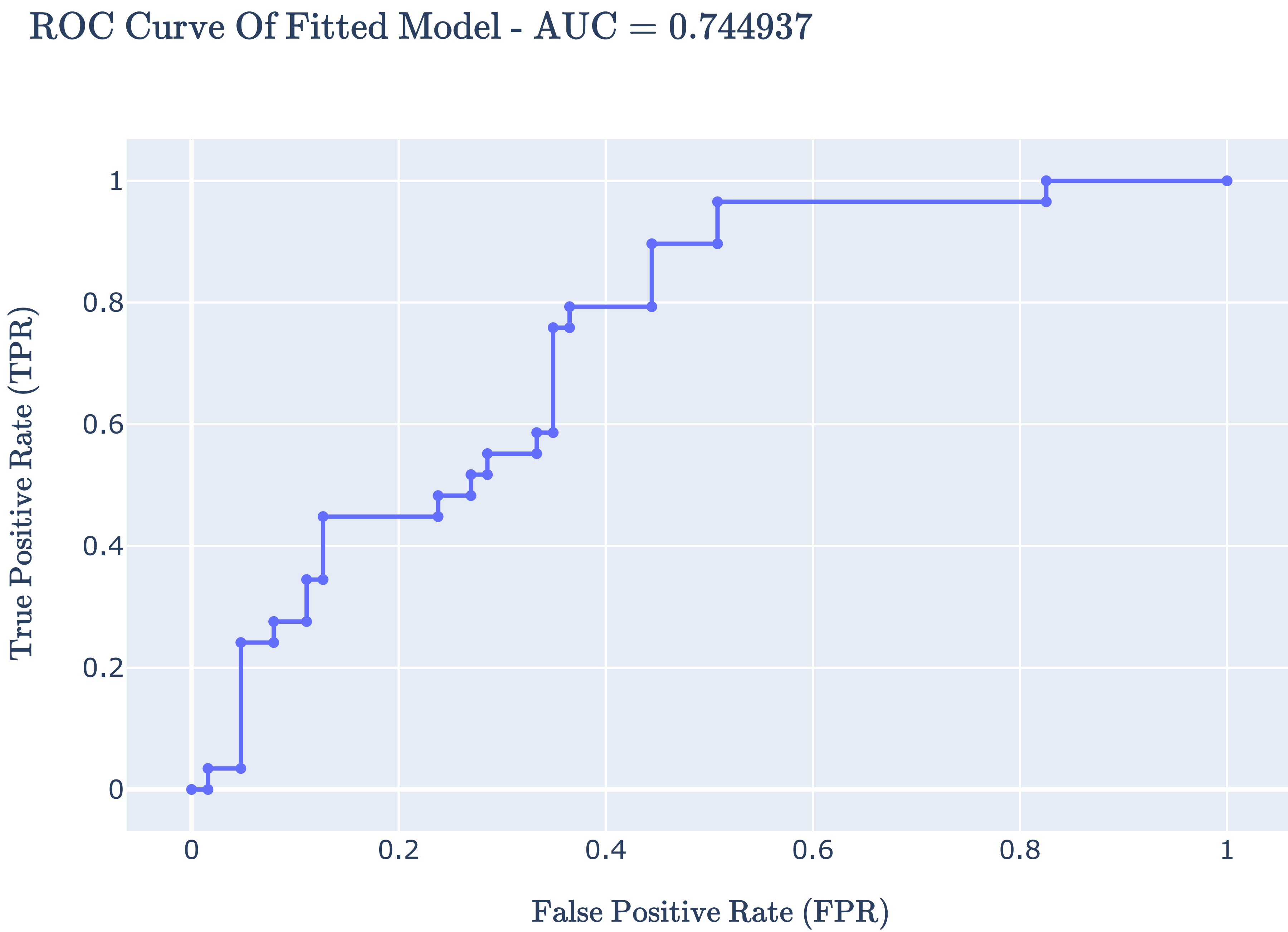

Convergence Rate: Norm L2, eta:1
L2 Norm
3.5
3
2.5
2
0
200
400
600
800
1000
Iterations


Convergence Rate: Norm L2, eta:0.1
L2 Norm
1.5
1
0.5
0
Iterations
0
10
20
30
40



Convergence Rate: Norm L2, eta:0.01
L2 Norm
2.5
2
1.5
1
0.5
0
0
100
200
300
400
Iterations


Convergence Rate: Norm L2, eta:0.001
L2 Norm
0
0.5
1
1.5
2
2.5
0
200
400
600
800
1000
Iterations


ROC Curve Of Fitted Model - AUC = 0.744937
True Positive Rate (TPR)
False Positive Rate (FPR)
0
0.2
0.4
0.6
0.8
1