תרגיל יבש 5 ־מבוא למערכות לומדות

אורי זהר ־ 057069502

25 בינואר 2023

שאלה 1

נתון $X_1,..,X_{10}$ מ"מ בלתי תלויים המתפלים $Uniform$ על הקטע $[0,\theta]$,
עבור $\theta > 0$ לא ידועה.

1.1

חשב את פונקציית הצפיפות עבור משתנה מקרי X_i:

$$P(X_i = x; \theta_i) = \frac{1}{\theta}\mathbb{I}\{x_i \geq 0, x_i \leq \theta\}$$

1.2

חשב את פונקציית ה־$Likelihood$:

$$L(X_1, ..., X_{10}; \theta) = P(X_1 = x_1, X_2 = x_2, ..., X_{10} = x_{10}; \theta) =^{iid}$$

$$\prod_{i=1}^{10} P(X_i = x_i; \theta) = \prod_{i=1}^{10} \frac{1}{\theta}\mathbb{I}\{x_i \geq 0, x_i \leq \theta\} = \frac{1}{\theta^{10}}\mathbb{I}\{\forall i \in [10] : x_i \geq 0, x_i \leq \theta\}$$

1.3

גזור את פונקציית ה־$Likelihood$ ובדוק איזה θ מקיים $\frac{\partial L}{\partial \theta} = 0$

$$\frac{\partial L}{\partial \theta} = \frac{\partial}{\partial \theta}(\frac{1}{\theta^{10}}\mathbb{I}\{x_i \geq 0, x_i \leq \theta\}) = \frac{-10}{\theta^{11}}\mathbb{I}\{\forall i \in [10] : x_i \geq 0, x_i \leq \theta\}$$

נשים לב שכאשר $\theta \longrightarrow \infty$ נקבל $\frac{-10}{\theta^{10}} = 0$ ולכן $\frac{\partial L}{\partial \theta} = 0$.

בנוסף לכל θ המקיימת : $0 < \theta < max\{x_i | i \in [10]\}$
נקבל $\mathbb{I}\{\forall i \in [10] : x_i \geq 0, x_i \leq \theta\} = 0$ ולכן $\frac{\partial L}{\partial \theta} = 0$.

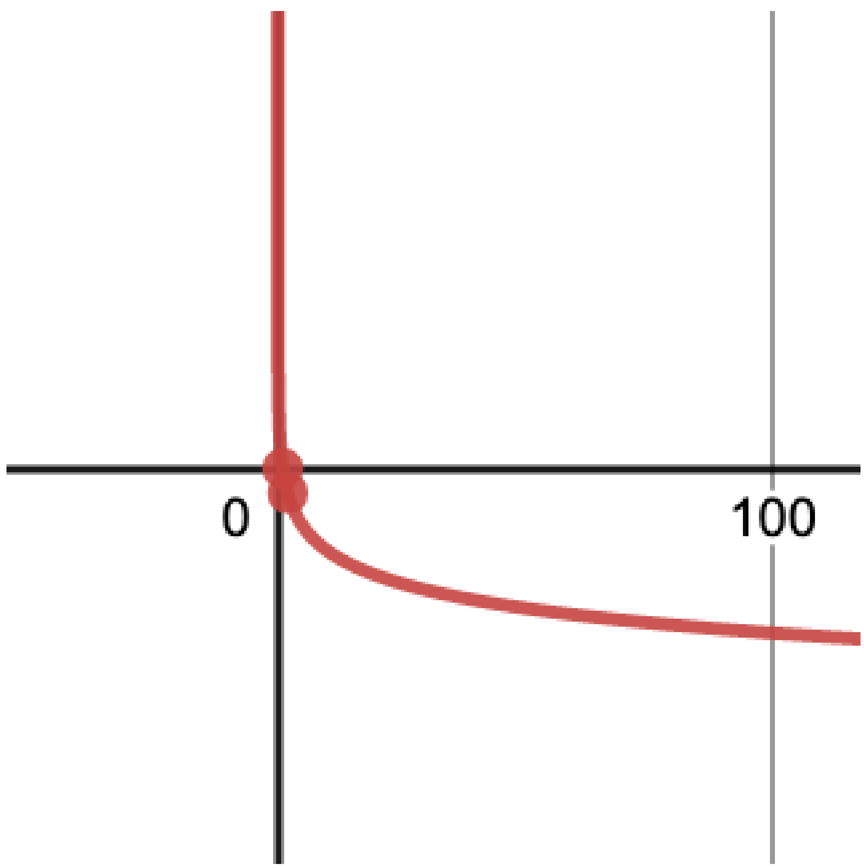
1.4

הצג את הגרף של $log(L(X_1, ..., X_{10}; \theta))$:

$$log(L(X_1, ..., X_{10}; \theta)) =$$

$$log(\frac{1}{\theta^{10}} \mathbb{I}\{\forall i \in [10] : x_i \geq 0, x_i \leq \theta\}) =$$

$$-10 log(\theta) \mathbb{I}\{\forall i \in [10] : x_i \geq 0, x_i \leq \theta\}$$



1.5

כפי שראינו בסעיף 3.1 אם קיים x_i כך ש $x_i < \theta$ אז נקבל שפונקציית ה$Likelihood$ מתאפסת לכן בהכרח נרצה לבחור θ כך שכל ה־x_i יהיו מוכלים בקטע $[0, \theta]$.
עבור בחירת θ כזאת נקבל ש־ $L(X_1, ..., X_{10}; \theta) = \frac{1}{\theta^{10}}$ שזוהי פונקצייה מונוטונית יורדת לכן נרצה גם שה־θ שנבחר היא המינימאלית האפשרית.
משני הטענות הנ"ל נקבל $\hat{\theta}_{MLE} = max\{x_i | i \in [10]\}$.

שאלה 2

2.1

נראה ש־ $F_{\alpha\theta} = cF_\theta$ עבור $c \in \mathbb{R}$.
נסמן

$$\hat{\theta} = (\hat{W}^{(1)}, \hat{W}^{(2)}, ..., \hat{W}^{(L)}), \hat{W}^{(i)} = \alpha W^{(i)}$$

$$\hat{h}^{(k)}(x) = \sigma(\hat{W}^{(k)^T} \hat{h}^{(k-1)}(x))$$

$$\hat{h}^{(1)}(x) = \sigma(\hat{W}^{(1)^T} x)$$

כעת נמצא את ה־c שיקיים את הנדרש :

$$F_{\alpha\theta} = \hat{W}^{(L)} \hat{h}^{(L-1)}(x) =$$

$$\hat{W}^{(L)} \sigma(\hat{W}^{(L-1)^T} \hat{h}^{(L-2)}(x)) =$$

$$\alpha W^{(L)} \sigma(\alpha W^{(L-1)^T} \hat{h}^{(L-2)}(x)) =$$

$$\alpha^2 W^{(L)} \sigma(W^{(L-1)^T} \hat{h}^{(L-2)}(x)) =$$

$$... = \alpha^k W^{(L)} \sigma(W^{(L-1)^T} \hat{h}^{(L-2)}(x) ... \sigma(W^{(L-k)^T} \hat{h}^{(L-k)}(x)) =$$

על מנת להגיע לתנאי העצירה נדרוש :
$L - k = 1$ כלומר $k = L - 1$
ונקבל :

$$\alpha^{L-1} W^{(L)} \sigma(W^{(L-1)^T} \hat{h}^{(L-2)}(x) ... \sigma(W^{(L-k)^T} \hat{h}^{(1)}(x)) =$$

$$= \alpha^{L-1} W^{(L)} \sigma(W^{(L-1)^T} \hat{h}^{(L-2)}(x) ... \sigma(\hat{W}^{(1)^T} x)) =$$

$$\alpha^{L} W^{(L)} \sigma(W^{(L-1)^T} \hat{h}^{(L-2)}(x) ... \sigma(W^{(1)^T} x)) = \alpha^L F_\theta$$

לכן עבור $c = \alpha^L$ הטענה מתקיימת .

2.2

כאשר $\alpha \longrightarrow 0$
מתקיים

$$F_{\alpha\theta} = a^L F_\theta = 0$$

ולכן

$$\frac{1}{1 + e^{-F_{\alpha\theta}}} = \frac{1}{1 + e^0} = \frac{1}{2}.$$

בסהכ אנחנו מקבלים שהמודל מתכנס להסתברות 0.5.