

26/6/20 - 1

11 \Rightarrow k(3))

סטוכסטי גרידנט דסנסט | Stochastic Gradient Descent

INTRODUCTION

כטבנ כתו פקודה כטבנ נילגיניז'ה דנולע:

1. דענות מילא:

$$\min_w \frac{1}{m} \sum_i (y_i - \langle w, x_i \rangle)^2 : \text{Squared Loss} \quad \text{pg} \bullet$$

\$\rightarrow\$ ypa \$\leftarrow\$ regression w/ a squared loss

$$\min \frac{1}{m} \sum |y_i - \langle w, x_i \rangle| \quad : (L_1) \text{ Absolute Value Loss}$$

לעומת Squared Loss מושגנו על ידי קביעת אמצע המרחקים.

שאלה ג' היא מושג ייחודי. הטענה ש**יוניסטרים** הם קבוצה (אחת) loss (היבטים).

: Quadratic Programming .2

18.11.2016 תאריך פרסום: מילון מושגים וjęzykowe terminy (אנו שרים בתקופה).

Solvers ღ ის, რომელი აქტორის რჩევა შეგვარულია. (Support Vector Machine სტრუქტურა)

$$\min_w \|w\|^2 \text{ s.t. } \forall i y_i \langle w, x_i \rangle \geq 1 : \text{Hard SVM} \\ \text{vs} \\ \min_w (\|w\|^2 + L_2(\omega)) : \text{Soft SVM}$$

$$l^{\text{hinge}}(\omega, (x, y)) = \max_{\omega} \{0, 1 - y \langle \omega, x \rangle\}$$

• 3 جایزہ ایجاد

• גורניר בודק ניטות היבר, $\min_{\omega} \frac{1}{m} \sum_i \log(1 + \exp(-y_i \langle \omega, x_i \rangle))$

Ridge Regression .4

$$\min_{\omega} \frac{1}{m} \sum_i (y_i - (\omega, x_i))^2 + \lambda \|\omega\|_2^2$$

Lasso Regression .5

$$\min_w \frac{1}{m} \sum_i (y_i - \langle w, x_i \rangle)^2 + \lambda \|w\|_1$$

... רְגִינְטָלֶס יְמִינָה וּמִבְּנֵי רְגִינְטָלֶס מִתְּנִיסְתָּאַת אֲלֹמָנִים נְגִידָה וְעַד...

ללא פג' דנילר

$$= R^d - n \cup jk \text{ plus } =$$

2. бүркүлүп көңілдіктердің көмекшілігіндең тәсілдерін сипаттаңыз.

ר) (ל עונתקה ג. גזרנו : $\alpha \in [0, 1]$. ב) $u, v \in C$ ר. 10)

$$\alpha u + (1-\alpha)v \in C$$



• **סְנָאָת** 'רַב' ב-NU |כ|(נ) ו-ב,ו,ו כ-ב,א

পৰি f সকল দণ্ডের পথে C রিপ্লি, $f: C \rightarrow \mathbb{R}$ । তাহা দণ্ডের দৃশ্যমান 2

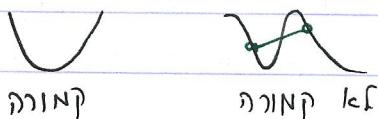
$\vdash \varphi \in [0, 1]$ if $\vdash \varphi$ and $\vdash \neg\varphi$ are both provable.

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

$\forall x: f''(x) > 0$: פlc גנור f Slc P"NYD גנור f: $\mathbb{R} \rightarrow \mathbb{R}$ plc *

R^d מ- \mathbb{R} קומפקט מ- $f(u) = \|u\|$

מִתְבָּאֵן כַּי תְּבָאֵן.



לְאַוִוָּה לְאַוִוָּה

$$\text{epigraph}(f) = \{(x, \beta) \mid f(x) \leq \beta\} \subseteq \mathbb{R}^{d+1}$$

for fine 13,pp 13,pp is very

ר' ר' $\text{epigraph}(f) \Leftrightarrow \text{וגונק ל} \in \mathbb{R}^d \rightarrow f$ גדרה א' : 286

\mathbb{R}^{d+1} -> $\mathcal{N}(\mu, \Sigma)$

בפרק (ג) נאמר כי רשותה מוסיפה לשליטה על הרים ונהר לא רק שטח אדמה, אלא גם רוח ורוחת הרים ונהר.

• 'Wif pinyin of lat ne 'feng

• **סינון** מינימלי של מילים בסיסי, בדרכו נסמן מילים שמייצגות מושגים נטולי ערך.

۱۴۰۷ - ۱۴۳۶ ﻦـ ﺖـ ﻚـ ﻞـ ﻪـ ﻢـ ﻮـ ﻢـ ﻪـ ﻮـ : ﻢـ ﻪـ ﻮـ ﻢـ ﻪـ ﻮـ ﻢـ ﻪـ ﻮـ

$B_r(u, r) = \{v \mid \|v - u\| \leq r\}$: 定義上， $B_r(u, r)$ 是一個以 u 為中心，半徑為 r 的球形鄰域。

• **SIC** . Ա-Ն Տարան կուղ ույլ . Տեղի բնույթ է

o s.t. $\forall v \in Br(u, r): f(v) \geq f(u)$

(ב) $v \in B(u)$ $\Leftrightarrow v - u \in B(0)$ $\Leftrightarrow \|v - u\| < r$

$$f(u) \leq f(u + \alpha(v-u)) \quad ; \rho, \gamma, N, \beta, f$$

$$f(u + \alpha(v-u)) = f(\alpha v + (1-\alpha)u) \leq \alpha f(v) + (1-\alpha)f(u) \stackrel{(*)}{\leq} f(v) : \text{由上不等式 } f$$

. 073) f α 1-նից բայց $f(u) \leq f(v)$: v ուղղի եթե $f(v) < f(u)$

$f(u) \geq f(w) + \langle \nabla f(w), u-w \rangle$: u ב- לכ גדרת רג'ולר f פlc : הוכחה

$$\cdot \text{ If } w \in \text{dom } f \text{ (e.g. } \mathbb{R}^n \text{)} \quad \nabla f(w) = \left(\frac{\partial f(w)}{\partial w_1}, \dots, \frac{\partial f(w)}{\partial w_n} \right) \quad \text{פוק:}$$

f יריר ב- \mathbb{R}^n . f מוגדר ב- \mathbb{R}^n ו- w נון-טורי ב- \mathbb{R}^n פkc ב- \mathbb{R}^n פkc ב- \mathbb{R}^n *

הוכחה:

: הוכחה של הטענה

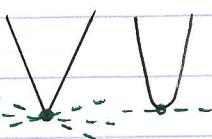
w יריר f יריר ב- \mathbb{R}^n subgradient v כפוק: הוכחה

$$f(u) \geq f(w) + \langle v, u-w \rangle \quad : u \text{ כפוק ב-} f$$

subgradients \rightarrow גדרת k ב- $\nabla f(w)$: הוכחה

, ב- \mathbb{R}^n $\nabla f(w)$ דק w -הגרדיאנט f פlc). w יריר f ב-

. (ב- \mathbb{R}^n w מינימום מקומי ב- f , ∞ ב- \mathbb{R}^n יריר f)



$\nabla f(w) \neq 0$: w כפוק ב- $f \Leftrightarrow$ f פkc ב- w :

$\Leftrightarrow (\exists o \in \nabla f(w)) \quad \text{w-הגרדיאנט של } f \text{ ב-} w$:

$\min_{\mathbb{R}^n} f \Leftrightarrow f \text{ ב-} w$:

$$\nabla f(w) \Leftrightarrow f(u) \geq f(w) + \langle \nabla f(w), u-w \rangle : u \text{ ב-} f \Leftrightarrow$$

$$\Leftrightarrow f(u) \geq f(w) : u \text{ ב-} f \Leftrightarrow f \text{ ב-} w$$

: $\forall w_1, w_2 \in \text{dom } f$ פlc $|f(w_1) - f(w_2)| \leq L \|w_1 - w_2\|$:

$$|f(w_1) - f(w_2)| \leq L \|w_1 - w_2\|$$

* נס. 3.1 \Leftarrow נס. 3.0.5-פ

$\forall w, \forall v \in \nabla f(w) : \|v\| \leq L \Leftrightarrow f$ פkc. מינימום f ב- w :

(L גודלו של f subgradients) (L גודלו של f פkc)

. f פkc. If v אונליין subgradient \rightarrow f פkc.

. f פkc. If v אונליין subgradient \rightarrow f פkc.

מינימיזציה דיסל

תכלית:

minimize $f_0(x)$

: $\mathbb{R}^d \rightarrow \mathbb{R}$ פונקציית מינימיזציה

subject to: $f_i(x) \leq b_i$ ($i \in [n]$)

, התחום הדרישה הוא $\text{dom}(f) \subseteq \mathbb{R}^d$ $\Rightarrow f: \mathbb{R}^d \rightarrow \mathbb{R}$ פונקציית מינימיזציה.

$f(\alpha x + (1-\alpha)z) \leq \alpha f(x) + (1-\alpha) f(z)$ פונקציית

$z, x \in \text{dom}(f), \alpha \in [0, 1]$

: $i \in [n]$ אם f_i מוגדרת כפונקציית מינימיזציה; גנריור מינימיזציה: אוסף כל פונקציית מינימיזציה f_i

: $i \in [n]$ אם f_i מוגדרת כפונקציית מינימיזציה: אוסף כל פונקציית מינימיזציה f_i

$\underset{\omega \in C}{\operatorname{argmin}} f(\omega)$ רכיבי אוסף מינימיזציה C מוגדרים כפונקציית מינימיזציה f .

(Half-Space) מוגדרת f : Feasibility Problem • מילוי
 $C = \mathbb{R}^d$: Unconstrained Minimization

Convex Learning Problems

השאלה היא האם ניתן למצוא מינימיזציה קווינית, כלומר כפונקציית מינימיזציה קווינית, \mathbb{R}^d על מנת למשוך נקודות על יישר אחד.

: מינימיזציה $v \in \mathbb{R}^d$ -ה מינימיזציה $h \in H$ בפונקציית l .
אלגוריתם (H, Z, l) מוגדר. $Z = X \times Y$ מינימיזציה כפונקציית מינימיזציה $l(\cdot, z)$: loss -> פונקציית מינימיזציה $l(x, \cdot)$ מינימיזציה כפונקציית $f_z(x) = l(x, z)$: $H \rightarrow \mathbb{R}$ פונקציית מינימיזציה f_z .

የኢትዮጵያ አስተዳደር, ፍዴራል, PAC-አንድነት ስንነሳ የሚከተሉ ቅጽ 5

PAC-ANSWER הצעת החלטה  LOSS → פגיעה •

• N_A / $\sum |f_i|$, ERM $\exists N \ni \forall \epsilon, \exists N$

וְנִזְמָנָה כַּנְדֵלֶת רְמִינָה וְלִזְמָנָה

⑩ High-speed network on PB

$\gamma = x \times y$, $(\#_1, \#_2)$ の組合せ : $N100-13001-120N1$ の組合せ

$f, B \in C_G(\Omega)$ pf (CLB) Convex-Lipschitz-Bounded λ $\in L^{\infty}$
 $: p, N \in W^{1,1}(\Omega)$ pf $p, q \in L^{\infty}$

$\|\omega\| \leq B : \omega \in H$ ԵՐ ՊՈԽՆԴ ԱՅլր ԽԸ Հ Ա 1

• מ.ב.ד.ס.פ-1! מ.מ.ר. פ.ר. א.ז., א.כ.ז, loss → פ.ר.א, z ∈ ℤ בד .�

רְגָנָרֶטִיבָּה וְלַearnable (רְגָנָרֶטִיבָּה) CLB → מְגֻנָּן ב- *

. γερά Ιζνή . ε, δ, ρ, Β

Stochastic Gradient Descent

Gradient Descent

טמי ומטבעה של רשות הרכבות הלאומית מושג עתה בראויים.

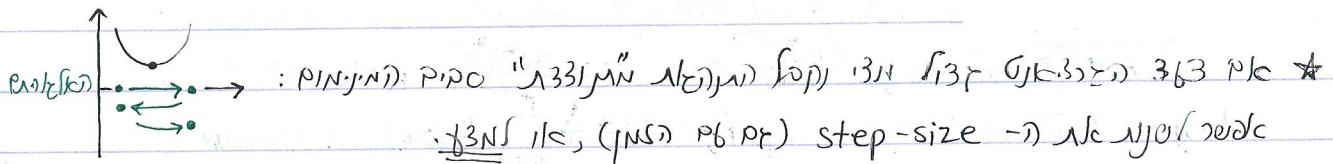
پاکستان

(אָמֵן) וְהַיְתָה מִלְחָמָה בְּנֵי יִשְׂרָאֵל בְּנֵי עֲמֹק .

$$\text{Step size } \eta \rightarrow 0$$

• አንቀጽ ሻጭ f (ለ ሃይል) ዘመኑን ተስፋይ ጥሩት ተከተሉ ነበርበታ የዚህ የዚህ ሻጭ

(உத்திர நிலைகள் மீது subgradient-ங்கள் என்றும் கூறலாம்)



: $P_{A \cup B}(\mathcal{C})$

Initialize: $\omega^{(1)} = 0$

$$\text{Update: } w^{(t+1)} = \underset{\mathcal{T}}{\text{arg}} \nabla f(w^{(t)})$$

$$\text{Output : } \vec{\omega} = \frac{1}{T} \sum_{t=1}^T \omega^{(t)}$$

1930-1941

સાહેબ નામની પત્રો

Feb 16 1952

: $\int_{\Gamma_C} \omega^{(f)} - \int_{\Gamma_D} \omega \mid_{\Gamma_C}$, גודלם ω : גודל Γ_C ו- Γ_D

$$f(w) \approx f(w^{(+)}) + \langle w - w^{(+)}, \nabla f(w^{(+)}) \rangle$$

76 נסיכי המלך הדריך אלס פג'ר פון פאלז יונילס ברגן סן. פס

: p(1), p(2) λ/c

$$\omega^{(t+1)} = \underset{\omega}{\operatorname{arg\,min}} \quad \frac{1}{2} \|\omega - \omega^{(t)}\|^2 + \eta (f(\omega^{(t)}) + \langle \omega - \omega^{(t)}, \nabla f(\omega^{(t)}) \rangle)$$

Overfitting of PNN \leftarrow PNN overfits the training set if Gradient Descent \star
PNN makes the prediction

Sub-Gradient Descent

• subgradient - $\nabla f(x)$ $\in \partial f(x)$

$v_t \in \mathcal{F}(w^{(t)})$ -> $w^{(t+1)} = w^{(t)} - \eta_{V,t}$: (1), update -> $\partial L \rightarrow NIS$

$w^* = \arg \min_w f(w)$ - ה ω המינימלית שפונקציית הערך $f(\omega)$ מושגת.

$\eta = \frac{\|w^*\|}{P\sqrt{T}}$: P.3f3 T ong Sub-Gradient Descent (W.3.3))

$$f(\bar{\omega}) \leq f_*(\omega^*) + \frac{\| \omega^* \|_P}{\sqrt{T}} \quad : \rho_{\bar{\omega} N} \underbrace{\bar{\omega}}_{\text{All } \bar{\omega} \in \Omega} \text{ (Eqn 7.6)} \quad \text{Eqn 7.6}$$

(תרכז)

շ 70 Եթե $\arg\min_w f(w) = w^*$ - ի ամենաքիչը լինի պահին էլ պահանջվում է պահանջմանը:

and find the Sub-Gradient Descent algorithm

(ω_{NN}) $\bar{\omega}$ Grön wörter sic, $\eta = \frac{\|\omega^*\|}{\rho\sqrt{T}}$ rf, N3, vlc

$$f(\bar{w}) \leq f(w^*) + \varepsilon \quad ; \rho \in \mathcal{P}_N$$

(subgradient) $\left[\begin{array}{c} \text{ಗ್ರಾಡಿಂ} \\ \text{subgradient} \end{array} \right]$

[Update - 1)]

$$\sum_{t=1}^T (f(w^{(t)}) - f(w^*)) \stackrel{(subgradient)}{\leq} \sum_{t=1}^T \langle w^{(t)} - w^*, v_t \rangle = \frac{\|w - w^*\|^2 - \|w - w^*\|^2}{2\eta} + \frac{\eta}{2} \sum_{t=1}^T \|v_t\|^2 : \text{?) DCIJ}$$

$$\frac{1}{T} \sum_{t=1}^T (f(w_t) - f(w^*)) \leq \underbrace{\|w^*\|_2 \gamma T}_{\mathcal{E}} + \frac{\eta \rho^2}{2} : \text{sk}, \quad t \leq T \quad \|V_t\| \leq \rho : \text{polynomial} - p-1 \quad \text{INP} \quad f$$

$$\mathcal{E} \geq R \cdot \epsilon \cdot \text{sk} \cdot \sqrt{\frac{\|w^*\|^2}{\gamma}} = \eta \quad T \geq \frac{\|w^*\|^2 \rho^2}{\eta^2} : \text{polynomial}$$

$$\cdot \varepsilon \geq R \Rightarrow \text{slc}, \sqrt{\frac{R w^*}{\rho^2 T}} = \eta, T \geq \frac{R w^* \rho^2}{\varepsilon^2} : \text{plc, } w^* \text{ is s.s.}$$

$$\text{.b) } \exists \rho \quad f(\bar{\omega}) - f(\omega^*) \leq \varepsilon \quad : \rho \int_1^T f(\bar{\omega}) \leq \frac{1}{T} \sum_{t=1}^T f(\omega_t) \quad : N, \text{INP}$$

$\|x_i\| = 1$ א"נ לכוד גודלו $\{(x_i, y_i)\}_{i=1}^m$ If $\exists \gamma > 0$ such that $y_i(\omega^\star, x_i) \geq \gamma$ for all i : linearly separable
 $y_i(\omega, x_i) > 0$: i בולו ω : found, גורם $y_i(\omega, x_i)$ linearly-separable (kind)
 $\gamma = \min_i y_i(\omega^\star, x_i)$: margin \rightarrow alc | no
 $\max_{\omega} \min_i y_i(\omega, x_i)$: margin \rightarrow alc optimization problem
 $\min_{\omega} \max_i -y_i(\omega, x_i)$: optimization problem גודל
 $f(\omega) = \max_i -y_i(\omega, x_i)$: function f ω optimization problem - alc \leftarrow
 $f(\omega) = \min_{\omega} f(\omega)$: function f ω optimization problem

\vec{w} argmax $-y_i \langle \vec{w}, \vec{x}_i \rangle - f$ \vec{w} is a \vec{w} - \vec{p} if f is subgradient.

Initialize: $\omega^{(1)} = 0$

النحو والصرف

For $t \in [\lceil \frac{1}{\gamma^2} \rceil + 1]$:

take $i \in \arg\max -y_i \langle w, x_i \rangle$

$$\text{update } \omega^{(t+1)} = \omega^{(t)} + \eta y_i x_i$$

$$\text{Output : } \bar{w} = \frac{1}{T} \sum_{t=1}^T w^{(t)}$$

$f(\omega^*) = -\gamma$: AN31977 πγγων πρι λεπτομέρεια προστασία προστασία προστασία

$f(\bar{w}) < f(w^*) + \gamma = 0$ (by N3) $\Rightarrow t > \frac{\gamma}{\eta_2}$ (N1)

3.20 N 110. N - M 100 W 155

Stochastic Gradient Descent

$$L_p(\omega) = \mathbb{E}[\ell(\omega, z)] : \text{寻求 } \min_{\omega \in \mathcal{H}} L_p(\omega) : \text{令 } \mathcal{H} \text{ 为可行集}$$

ERM ג'ז ג'זטס ל'ז, $L_D(w)$ alc הילן גוטמן סינ ויזטן ג'ז

$\exists D - \delta$ $\forall \ell(w_1, z)$ $\exists \gamma_{\ell w_1} : \text{SLC}$ ($\forall \ell$ $\forall j \in \{0, 1, 2\}$) L_j \in SLC

($\nabla \Pi^{\alpha\beta}$). Is ρ a $\partial_{\alpha}\Pi^{\alpha\beta}$ or $\partial_{\beta}\Pi^{\alpha\beta}$?

• (j) $\hat{\mu}_n$ is an unbiased estimate for the population mean μ . ★

Initialize: $\omega^{(1)} = 0$

: $P_{\lambda, \gamma, \beta, k}$

For $t \in [T]$:

choose $z_t \sim p$

let $v_t \in \partial L(\omega^{(t)}, z_t)$ update $\omega^{(t+1)} = \omega^{(t)} - \eta v_t$

Output: $\bar{\omega} = \frac{1}{T} \sum_{t=1}^T \omega^{(t)}$

pk ε 70 bf sic. p1B prounan r6 CLB λyP lɔr : 5N

$L_D(\omega)$ 18SNF Stochastic Gradient Descent p3, 1N ynjc

$$\eta = \sqrt{\frac{B^2}{P_T}} \quad P_61, T \geq \frac{B^2 P^2}{\varepsilon^2} : N(3) 0.1 \rightarrow \text{en r6}$$

: p"pN (p73INNA 1g(y)) $\bar{\omega}$ ofen sic

$$\mathbb{E}[L_D(\bar{\omega})] \leq \min_{\omega \in \mathcal{H}} L_D(\omega) + \varepsilon$$

- $L_D(\omega)$ 18SNF batch \leftarrow (nun)
- $L_D(\omega)$ 18SNF batch-learner
- deep learning -f pig foip