در اینجا [illegible]

$$w(t+1) = w(t) - \rho_t \left.\frac{\partial J(w)}{\partial w}\right|_{w=w(t)}$$

مشتق J در نقطه $w(t)$

نقاط اشتباه [illegible]

$$\frac{\partial J(w)}{\partial w} = \sum_{x \in Y} \delta_x x$$

در اینجا نقاطی که مشکوک هستند را در نظر گرفتیم؟

چرا اسمش : (کروس)

چون اما اینجا مشکل دیگر هست پس در نقاط درست هم معتبر خواهد بود

$$w(t+1) = w(t) - \rho_t \underbrace{\sum_{x \in Y} \delta_x x}_{\text{correction vector}}$$

اصلاح وزن‌ها

همه وزنها تکه به تکه

تا اینجا الگوریتم پرسپترون را معرفی کردیم.

در ابتدا وزنهای اولیه $w(0)$ تصادفی انتخاب می‌شود در هر گام با

تا جایی که به وزنهایی برسیم که مستقیم [illegible] یعنی الگوریتم همگرا بشه

مرحله به مرحله، میزان خطا کاهش می‌یابد از یک مقطعی به بعد به min

قضیه همگرایی پرسپترون (جهت وحدت کنیم)

به اندازه کافی دیگه نه

طبیعی + thermal loss

گرمش کن

درزبندیت

راه حل ب &

مواد سیم

آلومینیوم —

نه برنج (پلاستیک)

حصیر —

سرمه

Date: / / Subject:

[illegible]

[illegible]: $w(t+1) = w(t) - \rho_t \sum_{x\in Y} \delta_x x$

$\alpha > 0,\ \alpha \in R$
w^*: solution

[illegible]

$w(t+1) - \alpha w^* = w(t) - \alpha w^* - \rho_t \sum_{x\in Y}\delta_x x$

$$\|\vec w(t+1) - \alpha \vec w^*\|^2 = \|\vec w(t) - \alpha\vec w^*\|^2 + \rho_t^2\Big\|\sum_{x\in Y}\delta_x \vec x\Big\|^2$$

$$-2\rho_t\sum_{x\in Y}\delta_x\big(\vec w(t)-\alpha\vec w^*\big)^T x$$

میدانیم: $-\sum_{x\in Y}\delta_x \vec w_{(t)}^T \vec x < 0$

پس: $\|w(t+1)-\alpha w^*\|^2 \le \|w(t)-\alpha w^*\|^2 + \rho_t^2\Big\|\sum_{x\in Y}\delta_x \vec x\Big\|^2$

$$+2\rho_t\alpha\sum_{x\in Y}\delta_x \vec w^{*T} x$$

$\beta^2 = \max_{\tilde Y \subseteq \omega_1\cup\omega_2} \Big\|\sum_{x\in\tilde Y}\delta_x \vec x\Big\|^2$ [illegible]

$\gamma = \max_{\tilde Y\subseteq \omega_1\cup\omega_2}\sum_{x\in\tilde Y}\delta_x \vec w^{*T}\vec x$ [illegible]

نتیجه: $\|w(t+1)-\alpha w^*\|^2 \le \|w(t)-\alpha w^*\|^2 + \rho_t\beta^2 - 2\rho_t\alpha|\gamma|$

فرض کنید $\alpha = \frac{\beta^2}{2|\gamma|}$

پس $\|w(t+1)-\alpha w^*\|^2 \le \|w(0)-\alpha w^*\|^2 + \beta^2\Big(\sum_{k=0}^{t}\rho_k^2 - \sum_{k=0}^{t}\rho_k\Big)$

$t = t_0$

$\lim_{t\to\infty}\sum_{k=0}^{t}\rho_k = \infty$

$\lim_{t\to\infty}\sum_{k=0}^{t}\rho_k^2 < \infty$

[illegible]

$A \le 0$ [illegible]

پس $0 \le \|w(t_0+1)-\alpha w^*\| \le 0 \Rightarrow w(t_0+1) = \alpha w^*$

[illegible]

Date :

Subject :

توابع زیان ← مسائل طبقه‌بندی (کلاس‌بندی)

- آنتروپی
- ...
- Hinge

توابع زیان ← مسائل رگرسیون (کمّی، پیوسته)

- میانگین مربعات خطا
- میانگین قدر مطلق خطا
- تابع Huber
- تابع logcosh
- تابع چندک

۱. میانگین مربعات خطا : mean square error : MSE : L^2

$$MSE = \frac{\sum (y_i - \hat{y}_i)^2}{n}$$

حساس به داده پرت

۲. میانگین قدر مطلق خطا (mean absolute error) : MAE : L^1

$$MAE = \frac{\sum |y_i - \hat{y}_i|}{n}$$

داده پرت

۳. $\sqrt{MSE} = RMSE$ ریشه میانگین مربعات خطا

لذا استفاده از MSE و MAE ...

وقتی که داده پرت وجود داشته باشد ...

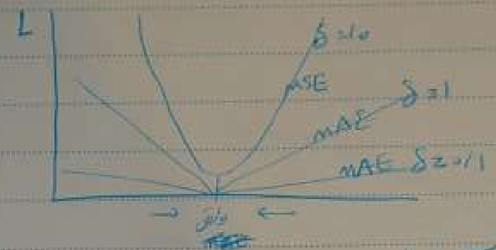
MAE < RMSE

Date: / / Subject:

تابع زیان هوبر — میانگین خطای قدر مطلق هموار شده Smooth mean Absolute error

مشتق پذیر است ، در برابر داده پرت تأثیر پذیر نیست ✓

$$L_\delta(y,\hat{y}) = \begin{cases} \frac{1}{2}(y-\hat{y})^2 = MSE & |y-\hat{y}| \le \delta \quad \text{ناحیه ۱} \\ \delta|y-\hat{y}| - \frac{1}{2}\delta^2 = MAE & |y-\hat{y}| > \delta \quad \text{ناحیه ۲} \end{cases}$$



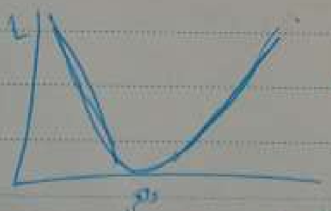
پس انتخاب δ مهم است !

تابع زیان لگاریتم کسینوس هذلولی (Log cosh)

cosh

logarithm of Hyperbolic cosine

$$L_{log\_cosh}(y,\hat{y}) = \sum_{i=1}^{n} \log(\cosh(y-\hat{y}))$$



log(cosh(x)) — مقادیر کوچک — $\frac{x^2}{2}$

log(cosh(x)) — مقادیر بزرگ — $|x| - \log(2)$

مشابه هوبر است ، مشتق دوم هم دارد ، در خطاهای [illegible]

گرادیان بیشتر (هسین) [illegible]

Date: / /

Subject:

[illegible] (Quantile loss [illegible]

[illegible]

[illegible]

MAE [illegible]

$$L_\gamma(y,\hat{y}) = \sum_{i\in\{i:\,y_i<\hat{y}_i\}} (1-\gamma)\,|y_i-\hat{y}_i| + \sum_{i\in\{i:\,y_i\geq\hat{y}_i\}} (\gamma)\,|y_i-\hat{y}_i|$$

[illegible]

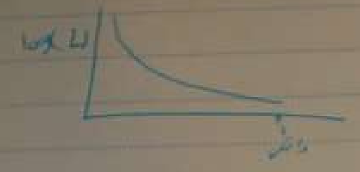
$\gamma = ۰/۲۵$

$\gamma = ۰/۵$

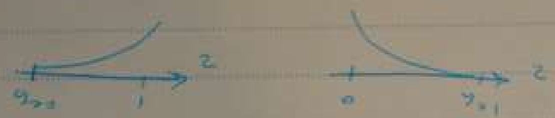
$\gamma = ۰/۷۵$

[illegible]

تابع CrossEntropy :

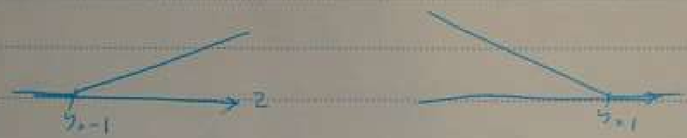


$-[y\log(z) + (1-y)\log(1-z)] \rightarrow$ رگرسیون لجستیک



خطای hinge : $\max(0, 1-yz)$

svm $\rightarrow$



خطای لجستیک : $\log(1+e^{-yz})$

رگرسیون لجستیک

چه در زیر مسئله چه در کل مسئله ...

N گروه ... تعدا

شرط ۲ آیا تابع تعمیم (m و D) و ...

$N \approx O(\frac{W}{G})$ غلط

ـه رخ خواهد داد بیش پارامتر ...

سوال ۱۱ M حالت با ۲ گروه (مستقل بشوند)

در نخستین شکل بالا ...

Date : / /　　　Subject :

بعد VC : بعد Vapnik : برای H [illegible] $VC(H)=d$ ، [illegible]

[illegible] با احتمال $1-\delta$ داریم :

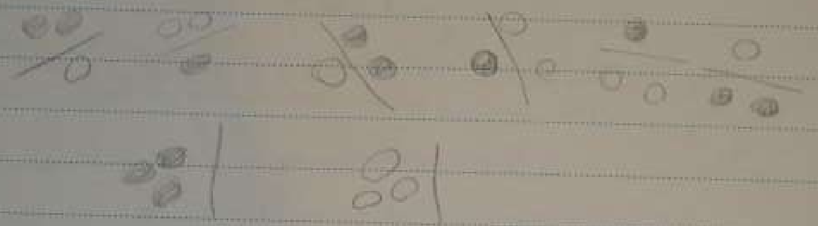
$$\varepsilon(\hat{h}) \le \left(\min_{h\in H} \varepsilon(h)\right) + O\left(\sqrt{\frac{d}{m}\log\frac{m}{d} + \frac{1}{m}\log\left(\frac{1}{\delta}\right)}\right)$$

VC ، [illegible] H [illegible] $VC(H)$ [illegible]

[illegible] H [illegible]

[illegible] H [illegible]

$VC(H) = 3$



[illegible] : اگر H [illegible] $|H|=k$ [illegible]

و δ ، m [illegible] با احتمال $1-\delta$ [illegible]

$$\varepsilon(\hat{h}) \le \left(\min_{h\in H} \varepsilon(h)\right) + 2\sqrt{\frac{1}{2m}\log\left(\frac{2k}{\delta}\right)}$$

[illegible] (Shattering) : [illegible] $S=\{x_1, \dots, x_m\}$ [illegible] H

[illegible] H [illegible] S [illegible]

[illegible]

$\exists h\in H ,\ \forall i\in[1,d] ,\ h(x_i)=y_i$

Date: / /

Subject:

[illegible] C [illegible]

[illegible]

حل تعریف [illegible] PAC [illegible]

[illegible] $[1-\delta]$ [illegible]

حل PAC [illegible]

[illegible] ϵ [illegible] $|H|$ [illegible]

[illegible] $\frac{1}{\epsilon}$ و $\frac{1}{\delta}$ [illegible]

[illegible] H [illegible]

[illegible] PAC [illegible] ϵ [illegible]

[illegible] hcH [illegible] $1-\delta$ [illegible]

[illegible]

$$m \geq \frac{1}{\epsilon}\left(\ln\left(\frac{1}{\delta}\right) + \ln|H|\right)$$

[illegible] PAC [illegible]

[illegible] C [illegible]

[illegible]

[illegible]

[illegible] agnostic [illegible] H [illegible]

[illegible] $1-\delta$

[illegible] ϵ [illegible] H [illegible]

$$m \geq \frac{1}{2\epsilon^2}\left(\ln\left(\frac{1}{\delta}\right) + \ln|H|\right)$$

[illegible]

[illegible]

Date: / /   Subject:

[illegible]

[illegible] H ← VC ← $\frac{VC(H)}{\delta}$

[illegible] H [illegible]

پس $m \geq \frac{1}{\varepsilon}\left(4\log_2\frac{2}{\delta} + 8\,VC(H)\log_2\frac{13}{\varepsilon}\right)$

$m \geq \max\left[\frac{1}{\varepsilon}\log\frac{1}{\delta},\ \frac{VC(H)-1}{32\varepsilon}\right]$

پس هیچ [illegible]

$VC \uparrow$ ←

$m = O\left(\frac{VC(H)+\log\frac{1}{\delta}}{\varepsilon}\right)$ با احتمال $1-\delta$

[illegible]

بهینه سازی تابع هزینه

فرض کنید θ پارامتر مجهول و $J(\theta)$ که هزینه مربوط است که باید min شود (مرحله‌های یادگیری)

1. الگوریتم گرادیان کاهشی
2. الگوریتم نیوتن
3. روش گرادیان مزدوج
4. بهینه سازی در مسائل دارای محدودیت
   1. محدودیت‌های تساوی
   2. محدودیت‌های نامساوی

شرط KKT (کاروش - کوهن - تاکر)

دوگان min-max

دوگان لاگرانژی

برنامه نویسی کانوکس

شکل دوآل ولف

Date: / /

Subject:

الگوریتم نیوتن: [illegible]

[illegible]

$$\Delta\theta = -H^{-1}(old)\,\frac{\partial S(\theta)}{\partial \theta}\bigg|_{\theta=\theta(old)}$$

[illegible]

$$= \frac{\partial S(\theta)}{\partial \theta}\bigg|_{\theta=\theta^0} + H(\theta-\theta^0)$$

پس $\theta_{new} = \theta_{old} - H^{-1}\left(H(\theta(old)-\theta^0)\right) = \theta^0$

پس min [illegible]

الگوریتم گرادیان مزدوج . [illegible]

$$\Delta\theta(t) = g(t) - \beta(t)\,\Delta\beta(t-1)$$

$$g(t) = \frac{\partial S(\theta)}{\partial \theta}\bigg|_{\theta=\theta(t)}$$

$$\beta(t) = \underbrace{\frac{g^T(t)\,g(t)}{g^T(t-1)\,g(t-1)}}_{\text{Fletcher-Reeves}} = \underbrace{\frac{g^T(t)\left(g(t)-g(t-1)\right)}{g^T(t-1)\,g(t-1)}}_{\text{Polak-Ribiere}}$$

Date: / /

Subject:

روش گرادیان نزولی: θ اولیه ← قدم

تکرار:

$$\theta_{(new)} = \theta_{(old)} + \Delta\theta$$

$$\Delta\theta = -\mu \left.\frac{\partial S(\theta)}{\partial \theta}\right|_{\theta=\theta_{(old)}}$$

μ: نرخ یادگیری

μ بزرگ ← نوسان و عدم همگرایی

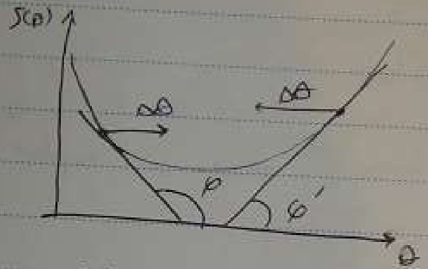
μ کوچک ← همگرایی دیر

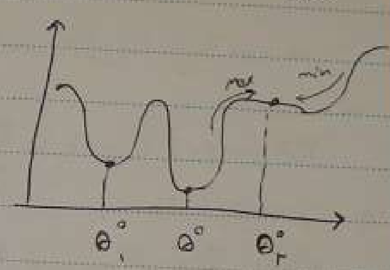
مناسب ← همگرایی به نقطه اکسترمم $S(\theta)$

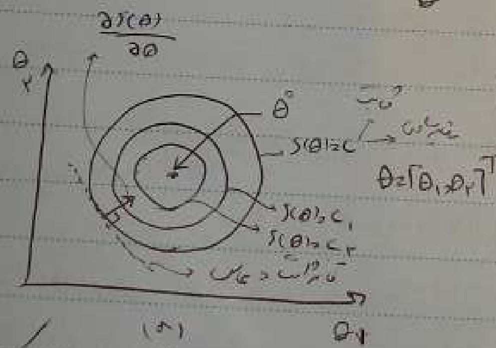
شیب = 0

$S(\theta)$: مینیمم کلی θ_1°، مینیمم محلی θ°، نقطه زینی θ_2°

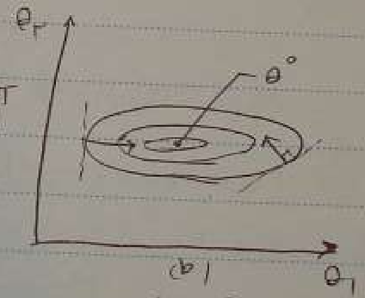
اما همگرایی به θ اولیه بستگی دارد ← به شکل $S(\theta)$







$\theta = [\theta_1, \theta_2]^T$



(الف) شکل دایره‌ای، نقطه شروع حرکت در جهت گرادیان، بهترین مسیر همگرایی بالاست. (مسیر مستقیم)

(ب) مسیر همگرایی از نقاط دورتر، حرکت در جهت گرادیان نیست. (مسیر نوسانی)

Date: / /

Subject:

$J(\theta)=c$ ، ثابت $\frac{\partial J(\theta)}{\partial\theta}$ در جهت بیشترین [illegible] $J(\theta)$

در هر نقطه یک سطح $J(\theta)=c$ داریم: $dc=0=\frac{\partial J(\theta)}{\partial\theta}^T d\theta \rightarrow \frac{\partial J(\theta)}{\partial\theta}\perp d\theta$

$$dJ=\frac{\partial J(\theta)}{\partial\theta}^T d\theta=\left|\frac{\partial J(\theta)}{\partial\theta}\right|\,\|d\theta\|\cos\phi$$

$\phi=90 \rightarrow \cos\phi=0$ min

$\phi=0 \rightarrow \cos\phi=1$ max

پس $\frac{\partial J(\theta)}{\partial\theta}$ [illegible] در جهت بیشترین کاهش $J(\theta)$ است $\Leftarrow$ $\frac{\partial J(\theta)}{\partial\theta}\parallel d\theta$

پس [illegible] در جهت [illegible] در [illegible] $J(\theta)$ حرکت کند.

[illegible] در [illegible] [illegible] θ [illegible]

$$J(\theta)=b-\vec{P}^T\vec{\theta}+\frac{1}{2}\vec{\theta}^T\vec{R}\vec{\theta}$$

سطح درجه ۲ ، فرض کنید:

$$\frac{\partial J(\theta)}{\partial\theta}=\vec{R}\vec{\theta}-\vec{P}=0 \longrightarrow \vec{R}\vec{\theta}=\vec{P}$$

پس $\theta(new)=\theta(old)+\Delta\theta$ ، $\Delta\theta=-\mu\left.\frac{\partial J(\theta)}{\partial\theta}\right|_{\theta=\theta(old)}$

$$\theta(t)=\theta(t-1)-\mu\left(\vec{R}\vec{\theta}(t-1)-\vec{P}\right)$$

(B) $\hat{\theta}(t)=\hat{\theta}(t-1)-\mu R\hat{\theta}(t-1)=(I-\mu R)\hat{\theta}(t-1)$

$\hat{\theta}(t)\equiv\theta(t)-\theta^\circ$

(A) $R=\Phi^T\Lambda\Phi$

ماتریس قطری

[illegible]

Date: Subject:

$\hat{\theta}(t) = Q\hat{\theta}(t)$

$\hat{\theta}(t) = \underbrace{(I-\mu\Lambda)}_{\text{ماتریس قطری}}\hat{\theta}(t-1)$

بردار ویژه $\hat{\theta}$ در θ

$\hat{\theta}_i(t) = (1-\mu\lambda_i)\hat{\theta}_i(t-1) \longrightarrow \hat{\theta} = [\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_L]^T$

به دلیل قطری بودن ماتریس معادلات از هم جدا می‌شوند

$\hat{\theta}_i(t) = (1-\mu\lambda_i)^t\hat{\theta}_i(0)$

اگر همگرا شود

$\lim_{t\to\infty}\hat{\theta}_i(t) = 0 \longrightarrow \lim_{t\to\infty}\theta_i(t) = \theta_i^*$, $i = 1, 2, \dots, L$

شرط همگرایی : $|1-\mu\lambda_i| < 1$, $i = 1, 2, \dots, L$

★ $\hat{\theta} \longrightarrow \theta^*$, if $\mu < \frac{2}{\lambda_{max}}$ پس ماتریس همگرا می‌شود

که بزرگترین مقدار ویژه R

R + مثبت +

پس سرعت همگرایی الگوریتم گرادیان کاهشی به نسبت $\lambda_{min}/\lambda_{max}$ کنترل می‌شود ★

• توابع درجه غیر درجه ۲:

بسط تیلور تقریب تیلور با ماتریس هسین

(ماتریس مشتقات جزئی درجه ۲) [مشتقات] مرتبه دوم

مقدار تقریبی که تابع [illegible] را دارد

(مشتقات جزئی تابع غیر درجه ۲، مرتبه دوم)

اگر $S(\theta)$ درجه ۲ نباشد، می‌توانیم از تقریب تیلور استفاده کنیم و فرض کنیم حول θ^* ...

θ^* ، $S(\theta)$ ... (به شکل درجه ۲)

$S(\theta) = S(\theta^*) + (\theta-\theta^*)^T g + \frac{1}{2}(\theta-\theta^*)^T H (\theta-\theta^*)$

این تقریب الگوریتم با ... در ماتریس هسین کنترل می‌شود

ماتریس هسین

$g = \frac{\partial S(\theta)}{\partial\theta}\Big|_{\theta=\theta^*}$

مشتق نسبت به θ^*

$H(i,j) = \frac{\partial^2 S(\theta)}{\partial\theta_i\partial\theta_j}\Big|_{\theta=\theta^*}$

ماتریس هسین نسبت به θ^*

Date: / /

Subject:

[illegible]

[illegible]

minimize $S(\theta)$

بشرط $A\theta = b$

[illegible]

hence A has full row rank

مرتبه سطری کامل ? → regularity assumption

روش‌های کار با مسئله چند کلاسه SVMs :

① روش درختی : [illegible]

~~[illegible]~~

muti class SVMS

OVR — [illegible]

K تا SVM [illegible]

[illegible]

[illegible]

OVO — [illegible]

[illegible]

max win

[illegible]

DAGSVM — [illegible]

SVM [illegible]

K [illegible]

[illegible]

weston & watkins — [illegible]

[illegible]

$n(k-1)$

(CS) crammer & singer

اما CS [illegible]