

(loss) of $\theta_0 \theta_1 \theta_2 \theta_3$ and $\theta_0 \theta_1 \theta_2 \theta_4$ leads to Δ_{loss} (loss)

and Δ_{loss} of $\theta_0 \theta_1 \theta_3 \theta_4$ leads to Δ_{loss} (loss)

لحاد تخفيف

weights of $\theta_0 \theta_1 \theta_2 \theta_3$ and $\theta_0 \theta_1 \theta_2 \theta_4$

leads to Δ_{loss} of $\theta_0 \theta_1 \theta_3 \theta_4$ and Δ_{loss} of $\theta_0 \theta_1 \theta_2 \theta_3$

لحاد تخفيف $\theta_0 \theta_1 \theta_2 \theta_3$ leads to Δ_{loss} of $\theta_0 \theta_1 \theta_2 \theta_4$

and Δ_{loss} of $\theta_0 \theta_1 \theta_3 \theta_4$ leads to Δ_{loss} of $\theta_0 \theta_1 \theta_2 \theta_3$

and Δ_{loss} of $\theta_0 \theta_1 \theta_2 \theta_4$ leads to Δ_{loss} of $\theta_0 \theta_1 \theta_2 \theta_3$

and Δ_{loss} of $\theta_0 \theta_1 \theta_3 \theta_4$ leads to Δ_{loss} of $\theta_0 \theta_1 \theta_2 \theta_3$

and Δ_{loss} of $\theta_0 \theta_1 \theta_2 \theta_3$ leads to Δ_{loss} of $\theta_0 \theta_1 \theta_2 \theta_4$

and Δ_{loss} of $\theta_0 \theta_1 \theta_2 \theta_3$ leads to Δ_{loss} of $\theta_0 \theta_1 \theta_3 \theta_4$

لحاد تخفيف $\theta_0 \theta_1 \theta_2 \theta_3$

لحاد تخفيف $\theta_0 \theta_1 \theta_2 \theta_3$ leads to Δ_{loss} of $\theta_0 \theta_1 \theta_2 \theta_4$

and Δ_{loss} of $\theta_0 \theta_1 \theta_2 \theta_4$ leads to Δ_{loss} of $\theta_0 \theta_1 \theta_3 \theta_4$

لحاد تخفيف $\theta_0 \theta_1 \theta_2 \theta_3$

contours plot $\theta_0 \theta_1 \theta_2 \theta_3$ and $\theta_0 \theta_1 \theta_2 \theta_4$

and $\theta_0 \theta_1 \theta_3 \theta_4$ leads to Δ_{loss} of $\theta_0 \theta_1 \theta_2 \theta_3$

and $\theta_0 \theta_1 \theta_2 \theta_3$ leads to Δ_{loss} of $\theta_0 \theta_1 \theta_2 \theta_4$

and $\theta_0 \theta_1 \theta_2 \theta_4$ leads to Δ_{loss} of $\theta_0 \theta_1 \theta_3 \theta_4$

لحاد تخفيف

contours plot $\theta_0 \theta_1 \theta_2 \theta_3$ and $\theta_0 \theta_1 \theta_2 \theta_4$

تَقْسِيمٌ طَارِئٌ وَّمُعَادٌ رُوَيْدَةُ بْنَ زَيْنٍ
أَرْجَعَهُ الْمُهَاجِرُونَ إِلَيْهِ مُرْتَفَعَهُ كَمَا يَرَى
كَمَا يَرَى الْمُهَاجِرُونَ إِلَيْهِ مُرْتَفَعَهُ كَمَا يَرَى
(MSE) *لِمَنْ يَرَى مُرْتَفَعًا فَلَمْ يَرَهُ فَلَمْ يَرَهُ*

۱۰) Minimum value of λ in order to get rid of outliers is
 ۱-۲٪ تفاوت این دو امتیاز در داده های ورودی و
 جمع همین درستگی λ که عبارت از $\lambda = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$ است

جیسے یہ ایک train ہے جو اس کا خوبی کرنے کے لئے ہے اور test set پر اپنے
اپنے ہی تکمیل کو کر کے خوبی کرنے کے لئے ہے اسی وجہ سے Overfitting
کا minimum Error کو دیکھ دیا جائے اور اس کے لئے یہ
کام و خود راستے کا جائز ہے اور اس کا وجہ یہ ہے
کہ اس کا یہی وجہ ہے کہ اس کا یہی وجہ ہے

the optimizer using \hat{L}_{opt} - is called cross-validated or leave-one-out optimizer.
• (Leave-one-out) gives a local model for the i^{th} data point
• generalization optimization (leave-one-out)

ویکی اسکرپت کیمی (DNN) میں ایسے overfitting کے نتائج
بھی ممکن ہیں جو مدل کو داداں کے مقابلے میں بھی خوبی کرنے کے لئے
کوئی حد تک رسائی کر جائے۔ اسی وجہ سے داداں کو داداں کے مقابلے میں
کچھ اضافی سیکھانے کے لئے اسے اپنے داداں کو داداں کے مقابلے میں
معین کرنا پڑتا ہے۔ اسی وجہ سے داداں کو داداں کے مقابلے میں
معین کرنے کا ایک طریقہ یہ ہے کہ داداں کو داداں کے مقابلے میں
معین کرنے کا ایک طریقہ یہ ہے کہ داداں کو داداں کے مقابلے میں

logistic regression کا جائزہ لے کر داداں کے مقابلے میں داداں کے مقابلے میں

logistic regression کا جائزہ لے کر داداں کے مقابلے میں داداں کے مقابلے میں

logistic regression کا جائزہ لے کر داداں کے مقابلے میں داداں کے مقابلے میں

logistic regression کا جائزہ لے کر داداں کے مقابلے میں داداں کے مقابلے میں

$$y = w_1 x_1^2 + w_2 x_2^2 + w_3 x_1 x_2 + b$$

graph

learning rate = 0.1

$b = -1$ \rightarrow initial value

beta = 0.9

$w_1 = +1$, $w_2 = -1$, $w_3 = -1$

batch size = 2

$$MSE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

x_1	x_2	y
1	-1	10
2	0	13
0	2	4
-1	1	4

(stochastic gradient descent) 2nd

فیصلہ دار رسمی

$$w_{\text{new}} = w_{\text{old}} - \eta \alpha \text{gradient}$$

first Batch $\Leftrightarrow 1 \text{ row}$

$$\hat{y}_1 = 1(1)^2 + 1(-1)^2 + 1(1)(-1) + 1 = 2$$

$$e_1 = 2 - 10 = -8$$

$$\hat{y}_2 = 1(2)^2 + 1(0)^2 + 1(2)(0) + 1 = 5$$

$$e_2 = 5 - 13 = -8$$

$$\frac{\partial}{\partial w} \sum e^2 \times \frac{\partial \hat{y}}{\partial w} \quad \leftarrow \begin{array}{l} \text{جواب میں ایک طرف} \\ \text{loss function} \end{array}$$

$\frac{\partial \text{loss}}{\partial w}$ سے ممکنہ ترین

$$\frac{\partial \text{loss}}{\partial w} \approx \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial w}$$

مسنون کوئی نہیں
لستہ وزن

$$\nabla w_1 = \frac{1}{2} (2(-8)(1)^2 + 2(-8)(2)^2) = -8 - 32 = -40$$

$$\nabla w_2 = \frac{1}{2} (2(-8)(-1)^2 + 2(-8)(0)^2) = -8 + 0 = -8$$

$$\nabla w_3 = \frac{1}{2} (2(-8)(1)(-1) + 2(-8)(2)(0)) = 8 + 0 = 8$$

$$\nabla b = \frac{1}{2} (2(-8)(1) + 2(-8)(1)) = -8 - 8 = -16$$

وزن کا درجہ

$$w_1 = 1 - 0.1(-40) = 1 + 4 = 5$$

$$w_2 = 1 - 0.1(-8) = 1 + 0.8 = 1.8$$

$$w_3 = 1 - 0.1(8) = 1 - 0.8 = -0.2$$

$$b = 1 - 0.1(-16) = 1 + 1.6 = 2.6$$

$$w_1 = 5 \quad w_2 = -0.2$$

$$w_3 = -1.8 \quad b = 2.6$$

for Batch $\Leftarrow 2$ No.

$$\hat{y}_1 = 5(0)^2 - 0.2(2)^2 - 1.8(0)(2) + 2.6 = 1.8$$

$$e_1 = 1.8 - 11 \Leftarrow -9.2$$

$$\hat{y}_2 = 5(-1)^2 - 0.2(1)^2 - 1.8(-1)(1) + 2.6 = 9.2$$

$$e_2 = 9.2 - 14 \Leftarrow 5.2$$

\Rightarrow Loss Function

$$\nabla w_1 = \frac{1}{2} (2(-9.2)(0^2) + 2(5.2)(-1)^2) = 0 + 5.2 = 5.2$$

$$\nabla w_2 = \frac{1}{2} (2(-9.2)(2^2) + 2(5.2)(1^2)) = -31.6$$

$$\nabla w_3 = \frac{1}{2} ((-9.2)(0)(2) + \cancel{2}(5.2)(-1)(1)) = -5.2$$

$$\nabla b = \frac{1}{2} ((-9.2)(1) + (5.2)(1)) = -4$$

\Rightarrow Loss Function

$$w_1 = 5 - 0.1(5.2) \Leftarrow 5 - 0.52 = 4.48$$

$$w_2 = -0.2 - 0.1(-31.6) \Leftarrow 2.96$$

$$w_3 = -1.8 - 0.1(-5.2) \Leftarrow -1.8 + 0.52 = -1.28$$

$$b = 2.6 - 0.1(-4) \Leftarrow 3.0$$

SGD + Momentum \Leftarrow

$$w_{\text{new}} = w_{\text{old}} - \beta \nabla_t U_t = \beta U_{t+1} + (1-\beta) \nabla w$$

standard SGD for loss function

$$\nabla w = (-30, -8, 8, -16)$$

$$w_1 = 1 - 0.1(-4) = 1.4$$

(epoch 0) \rightarrow چند چیزی را

$$w_2 = 1 - 0.1(-0.8) = 0.92$$

$$w_3 = 1 - 0.1(0.8) = 0.88$$

$$b = 1 - 0.1(-1.6) = 1.16$$

(epoch 0) \rightarrow چند چیزی را

$$\hat{y}_3 = 1.4(0)^2 + 0.92(2)^2 + 0.8(0)(2) + 1.16 = 2.52$$

$$e_3 = -2.52 - 4 = -13.52$$

$$\hat{y}_4 = 1.4(-4)^2 + 0.92(1)^2 + 0.8(-1)(1) + 1.16 = 2.72$$

$$e_4 = 2.72 - 4 = -1.28$$

(epoch 1) \rightarrow چند چیزی را

$$\nabla w_1 = \frac{2}{2} ((-13.52)(0) + (-1.28)(1)) = -1.28$$

$$\nabla w_2 = \frac{2}{2} ((-13.52)(4) + (-1.28)(1)) = -55.36$$

$$\nabla w_3 = \frac{2}{2} ((-13.52)(0) + (-1.28)(-1)) = 1.28$$

$$\nabla b = \frac{2}{2} ((-13.52)(1) + (-1.28)(1)) = -14.8$$

(-4, -0.8, 0.8, -1.6) \rightarrow V_1 پرتو چیزی را $\rightarrow V_2$ پرتو

$$v_{w1} = 0.9(-4) + 0.1(-1.28) = -3.728$$

$$v_{w2} = 0.9(-0.8) + 0.1(-55.36) = -6.256$$

$$v_{w3} = 0.9(0.8) + 0.1(1.28) = 0.848$$

$$v_b = 0.9(-1.6) + 0.1(-14.8) = -2.92$$

(epoch 1) \rightarrow چند چیزی را

$$w_1 = 1.4 - 0.1(-3.728) = 1.7728$$

$$w_2 = -0.92 - 0.1(-6.256) = -0.2944$$

$$w_3 = -1.08 - 0.1(0.898) = -1.1648$$

$$b = 1.16 - 0.1(-2.96) = 1.452$$

٣- (د) وقوع نزف بادیتی (Underflow) که از حد زیر میگذرد (Underflow) یعنی فاصله بین دو عدد متوالی کمتر از واحد عددی است.

(Overflow) از روی دسترسی در یک جایی بیش از حد ممکن (overshootting) میباشد. از روی دسترسی در یک جایی کمتر از حد ممکن (underflow) میباشد.

اما لذتی از آن نه (Divergence) از آن که در هر دو طرف از مقدار میگذرد.

از آنکه این دو نوع از خطاها را در یک مقدار میگذرد، باعث شدن از آنکه در هر دو طرف از مقدار میگذرد.

و در نتیجه نزف بادیتی (Underflow) میگذرد.

و در نتیجه نزف بادیتی (Underflow) میگذرد.

و در نتیجه نزف بادیتی (Underflow) میگذرد.

٤- (د) وقوع نزف بادیتی (Underflow) که از حد زیر میگذرد (Underflow) یعنی فاصله بین دو عدد متوالی کمتر از واحد عددی است.

و در نتیجه نزف بادیتی (Underflow) میگذرد.

(saddle)

۱/۲ نیکه زینو های در فریبا و میانی که ایستاده اند
بینهای ترازوی دارند (اینها نیز ممکن است در یک سایه باشند)

در چهار گونه روش مدار (مکانیزم)

در چهار طبقه مصنف اولیه های روش مدار (مکانیزم)

۱- ایستاده ایستاده ایستاده ایستاده

: SGD

اگرچه SGD آنچه برای اینجا مورد نظر نیست که باید در آنها در نظر گیری شود.

نیز که ایستاده (loss) در نظر می شود آنچه وزن

خوبی ایستاده (loss) را در دستور کنترل کنید که در آنها در نظر گیری شود.

نیزه الگوهای (Batch) (الجایزه) می شود

نیزه های از داده های ممکن است ممکن باشند که در آنها در نظر گیری شود.

که اینها ممکن است و ممکن نباشند

برای آنها ممکن است و ممکن نباشند

برای آنها ممکن است و ممکن نباشند

نمکیت هایی که در آنها ممکن است و ممکن نباشند

آنها ممکن است و ممکن نباشند

ادام (Adam) = Adaptive learning rate

ادام در نظر گیری که ایستاده های خوبی که در آنها در نظر گیری شوند

یا تغییرات بزرگ ایستاده های خوبی که در آنها در نظر گیری شوند

و یا تغییرات که در آنها در نظر گیری شوند

در آنها در نظر گیری شوند

جیسا کوئی سلسلہ کا ایک جزو ہے تو اسے smooth جیسا کہ اس کا مطلب ہے۔

دلتانیہ (Noise) اسے

Ch. Sug. Cylindrical & Ditch gradient \rightarrow very smooth flow
descent

وَيُنْهَا إِلَى الْمَسْكَنِ الْمُنْبَقِطِ وَالْمُنْتَهَى إِلَيْهِ بِالْمُرْسَلِينَ

گرامی خواسته شد که این واقعیت و حقیقیت کلی بتواند داده شود.

لـ ٦٤٣ - مـ ٢٠١٩ - تـ ٢٠١٩ - مـ ٢٠١٩ - لـ ٦٤٣

Eigenvalues) mini Batch gradient Descent Pages Noisy 1/21

ج (ع) (ع) (ع) (ع) (ع) (ع) (ع) (ع) (ع)

این نیز مجهود کوچک آنهاست از گردان و پخت است و در این

مکانیزم انتخابیت کنگره ایالتی

$$w^{t+1} = w^t - \eta \nabla J(w^t)$$

$$J(w^{t+1}) \leq J(w^t)$$

بیان ۷۰۵۰ کوچک:

$$f(x + \Delta x) \leq f(x), \quad \Delta x = -\eta \nabla f(x) \quad (\text{Bewegungsrichtung})$$

$$f(x + \Delta x) \approx f(x) + \nabla f(x)^T \Delta x$$

لـ ω كـ ω \Rightarrow $\omega = \omega$ \Rightarrow ω \in ω \Rightarrow ω \in ω

وَلِلَّهِ الْحُكْمُ وَإِلَيْهِ الْمُنْتَهَىٰ فَلَا يُنَزَّلُ بِأَنْوَافِهِ

$$J(w^{t+1}) \approx J(w^t) + (\nabla J'(w^t))^T \cdot \underbrace{(w^{t+1} - w^t)}_{\Delta w}$$

يمكننا استخدام
هذا لتقدير خطى

من دلایل در کارهای میان نویں تغییر وزن (برابر است با):

$$w^{t+1} - w^t = -\eta \nabla J(w^t)$$

جوابیگاری در شبکه های عصبی:

$$J(w^{t+1}) \approx J(w^t) + (\nabla J(w^t))^T \cdot (-\eta \nabla J(w^t))$$

برای ساده تر شدن این را در حالت متوسط می کنیم

$$\nabla^2 J(w^t) \cdot (-\eta \nabla J(w^t)) = -\eta \|\nabla J(w^t)\|_2^2$$

بنابراین

$$J(w^{t+1}) \approx J(w^t) - \eta \|\nabla J(w^t)\|_2^2$$

اینجا
 داشتیم
 و می خواهیم
 $\eta > 0$
 باشد

بنابراین این می تواند همانند مذکور شده باشد

$$J(w^{t+1}) \approx J(w^t) - (\text{a positive value}) \rightarrow J(w^{t+1}) \leq J(w^t)$$

ثابت ۲۰۵

SGD (stochastic) gradient descent کاریکاری Gradien Adam ...

sparse و چند جمله ای ... اینها ... High dimensional

... the reason is non-stationary ... it's not a function

... High order ... چند جمله ای ... ۱, ۲, ۳, ۴, ...

لاریم سیموون گیرن دا بولو ۱۶۰۰ و پروجیکت
لاریم سیموون گیرن دا بولو ۱۶۰۰ و پروجیکت
لاریم سیموون گیرن دا بولو ۱۶۰۰ و پروجیکت

~~دیکشنری مکالمہ~~ (۱)

Exponential Moving Average

$$g_t \leftarrow \nabla_{\theta} f_t(\theta_{t-1}) \quad \text{جواب مسأله ۲}$$

$$m_f = \beta_1 \cdot m_{f-1} + (1-\beta_1) \cdot (نیافرین تا دلخواه) \cdot \text{فنا و قتل}$$

$$\hat{m}_g \leftarrow m_g / (1 - \beta_1^t) \quad \text{Older molecules are discarded}$$

$$\hat{G}_g \leftarrow U_g / (1 - \beta_g^t) \quad \text{egs} \sim \text{...} \quad n$$

$$\hat{D}_f \leftarrow D_{f-1} - \alpha \cdot \hat{m}_f / (\sqrt{\hat{v}_f} + \epsilon) \quad \text{Durchsichtiger}$$

(and the other two are in the first part)

١٢-٥) مکانیزم و میداری

(ii) جنوبی پولیس اور پولیسیں کو درج کر دیں (M₁=0, M₂=0)

تمام (M₁=0, M₂=0) پولیسیں کو درج کر دیں (M₁=0, M₂=0)

اولیاً پولیسیں (B₁ & B₂) کو درج کر دیں اور اسے ایسا کہاں دیں کہ

اکٹلی پولیسیں اولیاً پولیسیں کو درج کر دیں اور اسے ایسا کہاں دیں کہ

اکٹلی پولیسیں (B₁ & B₂) RMSProf کے ساتھ پڑھ دیں

اکٹلی پولیسیں RMSProf کے ساتھ پڑھ دیں

اکٹلی RMSProf کے ساتھ پڑھ دیں

اکٹلی پولیسیں اولیاً پولیسیں کو درج کر دیں اسے ایسا کہاں دیں کہ

اکٹلی RMSProf کے ساتھ پڑھ دیں اسے ایسا کہاں دیں کہ

اکٹلی RMSProf کے ساتھ پڑھ دیں اسے ایسا کہاں دیں کہ

اکٹلی RMSProf کے ساتھ پڑھ دیں اسے ایسا کہاں دیں کہ

اکٹلی RMSProf کے ساتھ پڑھ دیں

SLD + momentum

(iii) ایک جگہ پولیسیں اولیاً پولیسیں کو درج کر دیں اسے ایسا کہاں دیں کہ

اکٹلی RMSProf کے ساتھ پڑھ دیں

اکٹلی RMSProf کے ساتھ پڑھ دیں اسے ایسا کہاں دیں کہ

اکٹلی RMSProf کے ساتھ پڑھ دیں اسے ایسا کہاں دیں کہ

اکٹلی RMSProf کے ساتھ پڑھ دیں

Adagrad (\rightarrow)

وہ sparse چیز کو کم پر جو گھٹے اور adam

(Vg)۔ اس کا وجہ Adagrad

کے لئے اس کا سبک $(\sum g^2)$ ہے جو اس کا وجہ Adagrad

اور adam کا وجہ میں فرق ہے اور اس کا وجہ

یہ اس کا وجہ ہے کہ adam کا وہ

لئے اس کا وجہ ہے اس کا وجہ ہے

RMSProp (\ominus)

لیکن RMSProp کا وجہ adam

کا وجہ RMSProp کا وجہ adam

کا وجہ adam کا وجہ ہے اس کا وجہ

کے وجہ ہے اس کا وجہ adam کا وجہ

کے وجہ ہے اس کا وجہ adam کا وجہ

(Iter 6 Eni کا) نیچے 3-5

لیکن SGD اور adam کا

کوئی دوسرے (2nd ch) MNIST کا وجہ (MP)

کوئی adam، CIFAR-10 کا وجہ (CVA) کا وجہ

کے وجہ ہے momentum کا وجہ

RMSProp کا وجہ SGD اور adam کا وجہ

لیکن momentum! SGD اور adam، then MNIST کا

وجہ ہے Adagrad کا وجہ

CNA (Adam) : $\theta_{t+1} = \theta_t - \eta_t \nabla J(\theta_t)$ Adagrad

و $\eta_t = \frac{\eta}{\sqrt{J'(\theta_t)}}$ $\theta_{t+1} = \theta_t - \eta_t \nabla J(\theta_t)$ adaGrad

و $J'(\theta_t)$ دوم در تابع دوگان دوام

adaGrad

5-5 Optimizations

نحوه و دستورات تابعیتی که میتوانند سریعاً از
 $(\nabla J(\theta))$ برای این تابع همچو تابع دوگان دوام

کار آمد و کمتر

(1) Adagrad: $\theta_{t+1} = \theta_t - \eta_t \nabla J(\theta_t)$ علت این کار در گرادیان

گردیده اند RMSProp که اینها را در گرادیان

(2) Automatic Annealing عیوب و مزایا

(initialization) از پیش بگیر

قاطعه تغییر و مسلسل

(intuitively interpreted) آنچه باز میگیریم

(invariant to rescaling) پس از دریافت تغییر مقیاس گرادیان

جایزه ای از اینجا (Regret bound) کار برای اینجا

$$O(\sqrt{T})$$

و اینجا نیز میگیریم که که اینجا

(Non convex) که اینجا

کار برای اینجا

Adam و Adam-Avg که اینجا

که اینجا که اینجا

الدروس والواجبات

1/ Notebook 1

الدروس والواجبات - Experiment 1

الدروس والواجبات - Full batch gradient descent

Learning rate = 0.001, 0.01, 0.1, 0.5, 0.8, 1.0

= 10 times

(136.65 درجات حرارة بخاري) = Full batch (1.00)

(172.39 درجات حرارة بخاري) = Mini batch (1)

Learning rate = 0.1, 0.5, 0.8, 0.9, 1.0

= 10 times (172.4 درجات حرارة بخاري)

2/ Full batch gradient descent

? 1 epoch = 100 درجات حرارة بخاري

الآن نريد أن نصل إلى درجة حرارة بخاري متساوية

الآن نريد أن نصل إلى درجة حرارة بخاري متساوية

لذلك

Full batch (100 درجات حرارة بخاري) = 1 epoch

Mini batch (50 درجات حرارة بخاري) = 2 epochs

Mini batch (25 درجات حرارة بخاري) = 4 epochs

ضرر کیا جائے گا آئندہ نتائج کو روپرکار
وزیر اعظم کا اعلان کیا تھا کہ اسی epoch میں سرکیمیت کو دیکھو
کہ ایک additional regularization effect کی وجہ سے نتائج کی میزبانی کیا جائے گا۔

کوئی پاک ایجاد نہیں کر سکتی ہے کہ ادیان میں کوئی، ایسے نہیں
 فوارڈ شارٹ minimum value کی جس (Stochasticity)
 وہ (generalization) ہے کہ flat minimum کا
 داری

جاءت نتائج (١٣٢٦ جلو) mini batch training = نتائج
جيدة جداً فرار از تراویثی خوب است و ممکن است
ابعد منطق full batch بسیار کم باشد
Momentum اینجا در experiment 2

السرعة المطلوبة (Convergence Speed) \rightarrow عدد الخطوات \rightarrow عدد الارقام

(6 epochs) ~~only~~ GPR + Nestleau ; SGD + Momentum
(8 epochs) SGD + Nestleau ; SGD + Nestleau

(26) 26-1 Bus & jumbo \rightarrow one isle \rightarrow 360 (standard)

partijen en andere "Pionier" pagina's zijn nog niet te vinden.

This Chapter SGD + Momentum, RMSprop and

2-1780.23

جاذب، تاریخ، سرعت

مکانیزم در مکانیک فیزیک: Momentum (مومانٹ) = $m \cdot v$
 پلٹر (Plaster) کا ایک مکانیزم ہے جو سارے بندوقیں اور گولے کو پلٹر کر دیتا ہے۔
 اس کا ایک نتیجہ ہے کہ پلٹر کا سارے بندوقیں اور گولے کو
 (Figure 6, p. 178) (مومانٹ کا تعریف)

Momentum کا لفظ ایسی ہے: restion و Momentum

(restion) = restion restion

restion (Look ahead) سے اپنی restion کا سارے

جذب کا سارے Momentum کا سارے

(% 24.36) سے اسیں Restion 560

جذب کا سارے Momentum کا سارے

Chopta Chhat, region (145 sq.m) - 1000 m.s.n.m
in 1st Reg

Some religious groups (R.R.O.) in the state

Chilean Sweet Potato (Ipomoea batatas)

سبت ۱۵ مهر ۱۳۹۷ / ۲۰۱۸ سهیل ۵۶۰ + مامنتم ۴۰

الجهاز المقاوم لـ $(LR = 0.5)$ \Rightarrow $B_g = 0.5$

• الْجَنَاحُ الْمُبِينُ

SGD: همچنان روشی است و ممکن است در دسته (۱۴۰) حرجیه کنمکده باشد

سیستم انتقالی (LPSO) است.

Wilson's Phalarope Nesting & Molt

مکالمہ فتحار دینے والیں کیوں

(0.5) = مجموع زیادی وقت نهادن برای این زنگ است.

گرما کی روزگار میں اپنے مکان پر افزاں سرعت

این کتاب را بزرگ باشید و بخوبی بخوانید و فرموده

⇒ Plotting points of C.R. if $\Delta \neq 0$ (overshooting, ringing etc)

.....

اسکم 56D کا ادیاول کامپنی اسٹارڈ چون ٹافنکنگ

وَالْمُؤْمِنُونَ إِذْ يَرَوْنَهُمْ يَقُولُونَ إِنَّمَا يَعْمَلُونَ بِأَنفُسِهِمْ فَإِنَّمَا
يُنَذِّرُ الْمُؤْمِنِينَ

دیگر پیش از اینکه میتوانیم این را بخواهیم

دین و نیز مذکور شده، ممکن است این دو نظریه را در نظر بگیریم.

پنجهای سیار سریع ترین دارند. تنظیم مقسیم Hyperparameters از وگرانی ارائه شده است. شرط ۰.۶ بین اینها میانگین است. که در آن روزهای آغازین آنها بسیار زیاد است و در نهایت به صفر رسیده است.

✓ notebook 2

✓ experiment 1

$$\text{Learning Rate} = 0.001$$

نمودار نشان می‌کند

Epoch 2: loss = 1.85
Epoch 3: loss = 1.85

این نتایج با RMSprop مطابقت ندارند.

نمودار نشان می‌کند

Epoch 21: loss = 1.89.21
Epoch 22: loss = 1.89.59
Epoch 23: loss = 1.88.60

این نتایج با RMSprop مطابقت ندارند.

loss = 1.88.27

overfitting

loss = 1.88.27

loss = 1.88.27

جیزیٹ سسٹم (Train Loss) سوچیکھیں اور جیزیٹ اس پر، جن کے
RMSPE، جن کے Only یہیں ہے (Test Loss) کیا گیا
(مثلاً 0.56 اور 0.36)

1. Stochastic gradient descent: experiment with weight decay vs. learning rate. Compare Adam vs. Gradient descent (Notebook p. 56) it performs better. It also gives good patches (Overfitting). Early stopping and early weight decay vs. early stopping with weight decay.

Experiment 2

جواب سوالات

Englehardt, Adam et al., Robustness of Lab.

دراسته نیز ۱۳۸۷

~~PLATE 3~~ Geological Survey of India R.M.S.P.R.O.P. Illustrations of Geology of India

البيانات المنشورة في الجدول أدناه تشير إلى أن متوسط درجة الحرارة في مصر هو 25°C (std dev = 0.3).

سُقْنَى لَارِوْنِ عَمَلَكَرْدْ آخِرْ دَائِسَتْ

نحوه انتشار دین میان میهن: $\frac{85.21}{1.87} = 45.4$ درصد در تغیرات دینی میان میهن

Öğrenci no: 009 öğrencisi BSMBGF, adı: Merve Zülüm Özdemir
Çılgın ve Jel Test 2023 Çılgın (1,88 kg)

Chloroform, with Singly or in 1 hr. Adagard
لـ ١ هـ تـ ٣

— اس بڑا دن سے نہیں ملے تھے اور اسی دن سے اس کی حکومت کا شروع 17 جولائی 1858ء کو ہوا۔

• Colorless

ادام و پروتکل / experiments

(اکتوبر 2015) Beta 1 نتیجے اور -
• صاف

لاب کو سریع ترین مکاری باتا۔ Beta = 0.9، t_{min} = 0.18 دن

(epoch تک درج کر دیں) (دستی خود کر دیں)

1% 89.58 الگوریتم دستی خود کر دیں Beta = 0.95، t_{min} = 0.2 دن
لپٹ کر دیں

(3.1.3) نتیجے اور اسکے پس پردہ

0.9 نتیجے اور اسکے پس پردہ Beta = 0.9
کوچک کر دیں جو بزرگ استفادہ از تاریخی گذشتہ و گذائلہ فتحی ایجاد کر دیں
لفڑائیں اور 0.99 باتیں نہیں فتحی از حد مسموں کا انتہا پر

سونپے لیں گے

(3.1.3) Beta 2 نتیجے اور -
• صاف

t_{min} (0.999) ایجاد کر دیں اور اسکے پس پردہ Beta = 0.99

لپٹ کر دیں (1.89.63) اور اسکے پس پردہ Beta = 0.99
لپٹ کر دیں

لپٹ کر دیں (0.9999 (0.99))

(1.89.0 ملے)

لپٹ کر دیں (Beta = 0.9999) ایجاد کر دیں اور اسکے پس پردہ

لپٹ کر دیں (0.9) ایجاد کر دیں اور اسکے پس پردہ

تجربی وارپا نزدیک این میزان را بیشتر برآورده است اما نتایج این دو مطالعه
 (جعفریان کرمانی و آندریوس و پاولوس) که از ۱۰۰ دیگر مطالعه ای داشتند نتایجی
 مسروق نهادند اما نتایجی که درین مطالعه ای داشتند ممکن است که در این مطالعه
 (بایلارچ و مددی) epsilon از ۰.۰۵ تا ۰.۰۷ باشد

مطالعه ای داشتند که در این مطالعه ای epsilon از ۰.۰۲ تا ۰.۰۳ باشد

این پژوهش حیفه بیانی ملحوظ بود از نظر قسم و تعداد اطلاعاتی که در این
 مطالعه ای داشتند که همان نتایجی که در این مطالعه ای داشتند این پژوهش را
 ترجیح نمی‌دانند

Robustness of Adam algorithm with respect to α and β .
 For $\epsilon = 0.5$, $P_2 = 0.9$ and $P_3 = 0.95$.