

Subject: جذب و انتقال

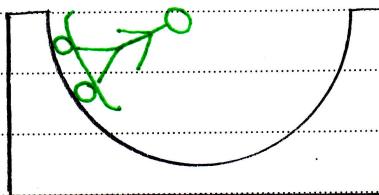
Year: Month: Day:

Momentum یا جو منتو ۳

ای از روش هریت حالت \hat{w} با دست اسقادران از Gradient Descent بمان

لیر مکان در \hat{w} هاست و موقتم بیانش جو منتو ۳ به ۴ اضافه شوند.

در سلسله زیر اسلیت سوار در $\min_{\hat{w}}$ استوار آن روی سود و بری تردد.



در هنریت و تحلیل تئوری دل فهم این موضوع (جو منتو ۳) وجود دارد.

در لینک های زیر نشان داده شده است (قابل GIF) اینجا جو منتو ۳ بیان از

لوگان همارد می شود. [جوابات](#)

gbhat.com/machine-learning/optimize_with_momentum.html

html

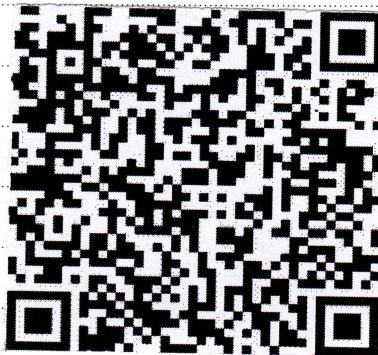
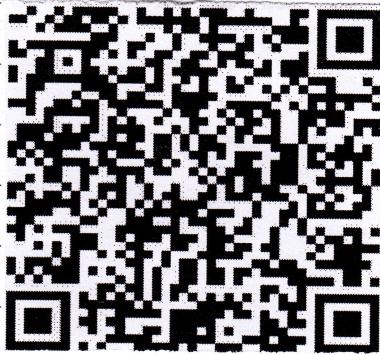
Kelar

www.kelarpooya.ir

I Subject: Year: Month: Day:

ghhat.com/machine-learning/gradient-descent-nesterov.html

nesterov.html



سوال دوستان

جیمینی ای پی ای دن "خطاب" \min از سبی استفاده نکنیم؟

ماعنی باشد سبی تغایر پذیری سدن لاست فا نلشون پی بهترم و از قبل هم املاعی

(راهن) باید ماداده نزدیک شود.

نمایشی ای میطل عقنا بخودرا explore (جاوش) لئنه عبارا اندیار نیاز داریم از لوطول

ها تاریخ سپم و در آنها لیر نکنیم.

I Subject: Year: Month: Day:

اپنے پڑائی کو سنبھالنے کا دستہ میں :=

$$w_{\text{new}} := w_{\text{old}} - \alpha \frac{\delta L}{\delta w_{\text{old}}}$$

Adding Momentum $\rightarrow w_{\text{new}} = w_{\text{old}} + m$

$$m = \beta m - \alpha \frac{\delta L}{\delta w_{\text{old}}}$$

$$\begin{matrix} \text{(new) } w \\ \uparrow \end{matrix} \quad \begin{matrix} \text{(old) } w \\ \uparrow \end{matrix} \quad \begin{matrix} \text{(old) } m \\ \uparrow \end{matrix}$$

$$\left\{ \begin{matrix} w^{t+1} := w^t + m^{t+1} \end{matrix} \right.$$

$$\left\{ \begin{matrix} m^{t+1} := \beta m^t - \alpha \frac{\delta L}{\delta w^t} \end{matrix} \right.$$

$$\text{initial } w^{t+1} \text{ لے جائیں}$$

سے خارج کرنے والے دادا کا نام m^{t+1} کیا ہے؟

$$\text{(new) } w^e = 1 , \beta = 0.9 , \alpha = 0.1 ; \delta L$$

1. Initialize $\rightarrow w^0 = 1 , m^0 = 0$

2. First ITR.

Kelar
www.kelarpooja.ir

Assume $\frac{\delta L}{\delta w} = 2$

$$m^1 = \beta m^0 - \alpha \frac{\delta L}{\delta w^0} = -0.2$$

Update $m^1 \rightarrow$

$$w^1 := w^0 + m^1 = 0.8$$

1 Subject:

Year:

Month:

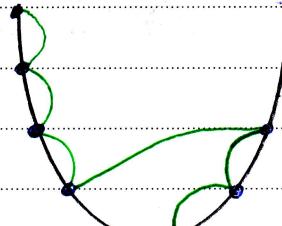
Day:

الا 1: اولین مجموعه صفر پی باش

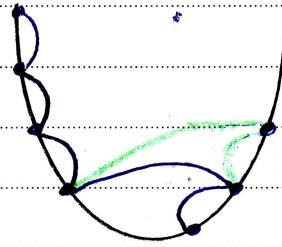
2: در فریم β_1 رای توان β در نقطه t رفت زیرا میل α (ریزیل دیت) مقدار

پلار، مکار (یعنی) بسیاری تغییرات داشته باشد

momentum



Nestrov momentum



در نوداری هست (راسی و عصبی) از نودار خود (نیزه) است و خوب معرف نیزه

بهینه تراست و معرف نیزه (نیزه) دارد. این اتفاق باعث شدن Nestrov (در

فارسی نیزوف) رخی (دید).

Nestrov momentum

$$\text{momentum} \rightarrow w_{t+1} := w_t + m^{t+1}$$

Nestrov
منبع جمع

$$m^{t+1} = \beta_1 m^t - \alpha \frac{\partial L}{\partial w^t}$$

Kelar

www.kelarpooja.ir

I Subject: Year: Month: Day:

$$\hat{w}^{t+1} = w^t + B_1 m^t$$

$$m^{t+1} = B_1 m^t - \alpha \frac{\partial L}{\partial w^t} (\hat{w}^{t+1})$$

$$w^{t+1} = w^t + m^{t+1}$$

نحوه (\hat{w}^{t+1}) با \hat{w}^{t+1} با توافق $\frac{\partial L}{\partial w^t}$ با توافق

$$w^0 = 1, \beta = 0.9, \alpha = 0.1 : J^0$$

1. Initialize: $w^0 = 1, m^0 = 0$

2. First ITR. $\hat{w}^1 := w^t + B_1 m^t = 1$

Assume $\frac{\partial L}{\partial w} (\hat{w}^1) = 2$ $m^1 := B_1 m^t - \alpha \frac{\partial L}{\partial w^t} (\hat{w}^{t+1}) = -0.2$

$$w^1 = w^t + m^{t+1} = 0.8$$

عمل درستار

يعني B_1 يعني ينحدر لـ w و m في بعده مبنية على نسخات امدادات حلول

پرس

Kelar

www.kelarpooja.ir

جاء ببريم ارجاع زياري با سنه خطي زيادي صوره دايره هر دفعه به داشت

1 Subject:

Year:

Month:

Day:

نحوه‌هون از β را بینجا (محض منتفع است) در جدول زیرشان (داده شده) باشند.

دستور لی	را دل) با هم منتفع	معتروف
نموده بروزرسانی	سرعت برآمدگانی تجربی از مرادیان معتبر پیش‌بینی شده Look ahead	سرعت برآمدگانی تجربی از مرادیان معتبر پیش‌بینی شده Look ahead
مزایا	فارغ‌فونوسنات در استای عمود بر جهت راهنمایی پیش‌بینی از مرادیان های آینده بهره‌مند می‌باشد	فارغ‌فونوسنات در استای عمود بر جهت راهنمایی پیش‌بینی از مرادیان های آینده بهره‌مند می‌باشد
بعایب	میان است برآمدگانی در min هایی ترین نیاز داشته باشد	سرعت برآمدگانی ممتاز ندارد
سرعت محمد رای	سریع راز مرادیان نزدیک ساده	سریع راز مرادیان نزدیک ساده
پیچیدگی ماساچوست	متوسط	متوسط
بارا مرها	نرخ پردازی (η) معنوی (β) و مومنتو (β)	نرخ پردازی (η) معنوی (β) و مومنتو (β)

۷. (ایتا) در جدول بالا کهان ۲۰ بی پاسد.

بارا مرها

بارا مرها معاذر داخلي بدلیل پارهای ماسن هستند در طی فرود نیز افزونی دارند از زاده ها

پاره قوه (تفاهم) بی سوون این بارا مرها محدود وزن دهانی باشند معاذر دارند های

نشیعی) یا غرایب در محل دهانی مدرسوون که می‌باشد بارا مرها می‌ستیان از زاده که آنوزنی

استراحت بی سوون در معاذر دیگران بعده خود را تفاهم بی سوون تا دل بتواند عمل دهنده

رویداده دهانی آنوزنی داشته باشد

1 Subject: Year: Month: Day:

وزن های شبکه عصبی (Weights)

بايسن های شبکه عصبی (biases)

هاي پردازش Les

نهایه پردازهای معادله های محاسبه شده از خروجی اخیر (خوب) مدل تعیین ایستادن

بررسی فرمایید. همچنان و ساختار مدل تأثیری ندارد. های پردازهای معادله

آنرا (پسون) جست و جوی کنید. همانند درود سرچ با جست و جوی نیافری

نتایج میتواند تغییر درست های پردازهای توانسته تحریری بر عکس مدل داشته

باشد

مدل های از های پردازهای

هزار (learningrate)

number of neurons per layer (عدد نورون های در لایه)

(number of layers)

I Subject: Year: Month: Day:

۰ آنرا بقیه تعداد نمونه های از پیش از برخور رسانی کمتر مانع نیست.

۰ تعداد دفعه آنرا بزیست (batchsize)

۰ تعداد دوره های آنرا تعداد دوره های مورد نظر (numberofepochs).

(numberofepochs).

20 | | | | 100 , batch size = 20 : بقیه *

۰ معمولاً بقیه سایز رو توانی از 2 قاری درست

((بیکم لایر نیم))

[github.com/Mohammad-Ali-Malekzadeh/Deep-](https://github.com/Mohammad-Ali-Malekzadeh/Deep-Learning-Logger)

Learning-Logger

۰ در درست ۵۵٪ باشد