

①

(۲) در بهینه سازی شبکه های عصبی، روش ششم و نهم یادگیری در فاکتور مهم

در همگرایی و کارایی الگوریتم بهینه سازی هستند.

روش ششم یک تکنیک برای شتاب دادن به گرایان در جهت مینی داکتیم سازی است.

این که می تواند باعث افزایش سرعت همگرایی و جلوگیری از مسدود شدن شود. روش ششم

با استفاده از یک Moving Average روی گرایان مینی داکتیم را محاسبه و بهادیت در هر

مرحله از گرایان مینی داکتیم نیز محاسبه می شود.

نرخ یادگیری اندازه قدم بهینه ساز در هر تکرار است. نرخ یادگیری بالا ممکن است باعث

الگوریتم شود و نرخ یادگیری پایین ممکن است باعث کند شدن یادگیری شود.

بنابراین به طوری که بتوان گفت روش ششم و نهم یادگیری در هر تکرار الگوریتم به جهت تأثیر

دادند. روش ششم بالا (از زیاد گرایان مینی داکتیم) و نرخ یادگیری پایین باعث همگرایی کند می باشد.

در روش ششم پایین (از کم گرایان مینی داکتیم) و نرخ یادگیری بالا باعث همگرایی سریع اما نادرست

نمودار مینی داکتیم می شود. بنابراین تنظیم مناسب این پارامترها با آزمون و خطا امکان پذیر است.

جدید برنامه برای نرخ یادگیری یک تکنیک برای تنظیم کردن نرخ یادگیری هنگام برودت آموزش است. نقش آن

کاهش آهسته نرخ یادگیری هنگام آموزش است. تنظیم کردن به یادگیری می تواند از مشکلات معمول

بهینه سازی مانند واگرایی و یا مینی داکتیم جلوگیری کند. این متغیر برای برنامه برای نرخ یادگیری

موجود دارد مانند Cosine Annealing, Exponential Decay, Step Decay به عنوان مثال در برنامه برای مینی

نرخ یادگیری در هر تکرار به صورت خطی کاهش می یابد.

به طوری که هدف اصلی این برنامه برای کاهش به تدریج کارایی مدل با همگرایی سریعتر و دقیقتر است.

آغاز نوروز (تعطیل)

(ج) ایضاً در عدد رتبه، مقدار رخ بادگیر را، یک ضرب ثابت گویند و در عدد

ممنون باشم اگر مقدور بود به شرح و دیگری دو پیکر و ضربات پای در هر یک

در تکرار های صدی فرخ باواری در ۱۰۵۰ و ۱۰۵۵ و ۱۰۶۰ می گذارد

سازمان تامین اجتماعی در ترکیه به صورت یکسره اجناس را میفروشد

(ج) کی لڑائی مع برابر رہتا cosine anealing است و لڑائی ضعیف لڑائی نامہ رہی ہا، مقدار، نفع یا دگرگی

حاضری نامہ کیسے دی از یک مقدار اور ایک سیکس شروع ہو ورت کیسے دی تا یک مقدار میسٹرم کیسے

محبوبہ بی بی محمدہ شہار خانہ دہلی سے آئے ہیں اور ان کے ساتھ ایک لڑکی اور ایک لڑکا بھی ہیں۔

صورت اداسی باشد. این از این گلو، بگلو، سینه و ریه ها، بمرکز در قوای از میهنیم های ملی.

تشدق آن به کامیابی بیشتر در مقامی زمینی که می‌کنند.

م. SGO یک سیستم بهینه سازی برای پرودرهای یاد از برای مدل هنگام آموزش است که برای

مورد سانی با دارا ماد و مرطه، عقد از گوارین تابع مزین حلقی از نونه نقلی استناد کرده

به عبارت دیگر در روزهای باران، اثر گرایی مای تبلی در نظر گرفته می شود و همین با این

هر هفته آشنایی یک قدم بلندتر ما به روز می‌ماند. همچنین از یک نرنگ بازی کتاب در 560 استان و کشور

از طرف دیگر Adam با انگه رستم بهینه سازی تطبیقی است که سعی در تخمین موفقیتم دارد. این روش

در واقع ترکیب وادای مزیت های در روش $RMSProp$ و $AdaGrad$ است. در روش اول هر پارامتر دارای

فرغ با دیکری خصوصی است که این بیدر مدل در مسائل با گرایان های sparse می کند. در روشی هم از یک moving average

دوی گم اینان ما برای اثر دادن به آن استفاده را کرد. در روش Adam از مشتق های دوم تابع هزینه نیز استفاده

مکانی در بنای این آستانه از این روشن‌ها، در آن‌ها را به طور عمده بر روی رسانی می‌کنند که به نام درنه آستانه (تقدیر) یاد کنند.

بهدرین - Ada نت به Ssd مدل ۸ ترابایت هکراس لایه و عملکرد بهتری دارد. از طرفی Ada حاوی سربا اتریهای بیشتری دارد.

بابی نیازم منابع بیشتر برای آن است.

نوع لایه	منطقه لایه			ابعاد ورودی			ابعاد خروجی			بار انتقالی	بار انتقالی	بار انتقالی
	نوع لایه	ابعاد	نوع لایه	نوع لایه	ابعاد	نوع لایه	ابعاد					
C ₁	۱۰	۹×۹×۳	۰	۱	۳	۱۰	۲۵۶×۲۵۶	۱۰	۲۵۶×۲۵۶	۱۰	۲۵۶×۲۵۶	۱۰
Relu	-	-	-	-	-	-	۲۵۶×۲۵۶	۱۰	۲۵۶×۲۵۶	-	-	-
C _p	۲۰	۷×۷×۱۶	۰	۲	۱۶	۲۰	۱۲۱×۱۲۱	۲۰	۱۲۱×۱۲۱	۲۰	۱۲۱×۱۲۱	۲۰
C _p	۴۰	۵×۵×۲۰	۰	۱	۲۰	۴۰	۱۱۶×۱۱۶	۴۰	۱۱۶×۱۱۶	۴۰	۱۱۶×۱۱۶	۴۰
Relu	-	-	-	-	-	-	۱۱۶×۱۱۶	۴۰	۱۱۶×۱۱۶	-	-	-
Max Pooling	-	-	-	-	-	-	۱۱۶×۱۱۶	۴۰	۱۱۶×۱۱۶	-	-	-
C _s	۵۰	۳×۳×۱۶	۰	۲	۲۰	۵۰	۵۸×۵۸	۵۰	۵۸×۵۸	۵۰	۵۸×۵۸	۵۰
flatten	-	-	-	-	-	-	۵۸×۵۸	۵۰	۵۸×۵۸	-	-	-
FC	۱۰	-	-	-	-	-	۱۴۰۰×۱	۱	۱۴۰۰×۱	۱	۱۴۰۰×۱	۱

عید نوروز (تعطیل)

(ب) بله، عدم استناد از تدابیر فعال ساز بعد از لایحه‌های پیشینی، باعث می‌شود که

فَسَدَ نَتَانَدِ الْكِدَاهِ غَيْرُ فَهْلٍ بَيْنَ دُرُودٍ وَفِرْدَوْسٍ (یا دیگره) از آغازی که عسکر

سچی دیک لایہ ایک نگر غفلت است، بدون استفادہ از تدابیر فعال سازہ، فروبی

این لایه همچنین خطی فداوند بر بنای این می‌کرانه در عملکرد مدل تأثیر منفی داشته باشد

و قدرت باد لری آن را گاهی دهد.

بنای این عدم استناد از تابع مثال ساز بعد از لایه های C_2 و C_3 از آغازی که تدائمی مدل در یادگیری

المعطل سعيد، و غرضی را گامی می دهد، بافت عملکرد ضعیف تر شده می شود.

عید نوروز (تعطیل)

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}_{5 \times 1}$$

$$W = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}_{2 \times 1}$$

(۳)

$$Z = \text{conv}(X, W) = \begin{bmatrix} w_1 x_1 + w_2 x_2 \\ w_1 x_2 + w_2 x_3 \\ w_1 x_3 + w_2 x_4 \\ w_1 x_4 + w_2 x_5 \end{bmatrix}_{4 \times 1}$$

چند نکته: هر مقدار فردی در این لایه نقطه به دو همان از ورودی بهشتی دارد و نه همه همان های ورودی، بنابراین می توان گفت استانه از یک نلتر 2×1 بجای ورودی های 5×1 می تواند خامیت نلک بدون مقایه فردی را همراه داشته باشد.

نکته دوم: اتصالات مزایای زیادی از جمله کاهش تعداد پارامترهای نیاز به یادگیری که باعث آسودن سیستم و بهینه تر شدن و همچنین کاهش احتمال overfit شدن مدل می شود. از طرف دیگر نکته سوم: اتصالات می تواند به شبکه کمک کند که ورودی ها را با ابعاد بزرگتر بدون نیاز به تعداد پارامتر بیشتر و ابعاد بزرگتر که همین امر منجر به بهینه شدن از نظر محاسباتی شبکه های عمیق می شود.

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}_{5 \times 1}$$

$$W = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix}_{4 \times 1}$$

(۴)

$$Z = \text{conv}(X, W) = \begin{bmatrix} w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4 \\ w_1 x_2 + w_2 x_3 + w_3 x_4 + w_4 x_5 \end{bmatrix}_{2 \times 1}$$

نکته آخر: هر مقدار فردی در این لایه به هم همان از کل ۵ همان ورودی بهشتی دارد بنابراین می توان گفت استانه از یک نلتر 5×1 بجای ورودی های 5×1 خامیت نلک بدون مقایه فردی زیادی ندارد.

۵۲) (آ) زمان سازی دسته ای می تواند با کم کردن Internal covariant shift که زمانی

در بهبود عملکرد شبکه های عصبی داشته باشد. مشکلی که ذکر شد در آیدزین شبکه های عصبی عمیق

زمانی رخ می دهد که توزیع داده های ورودی در لایه از شبکه تغییر کند. این توزیع با به روز رسانی

وایزهای لایه قبل تغییر می کند. حال اگر توزیع داده های ورودی در لایه تغییر کند، شبکه باید با تغییر وایزهای

لایه با این تغییر تطبیق یابد که این امر باعث کندی همگرای مدل و عملکرد ضعیف آن روی داده های جدید می شود.

کنشک زمان سازی دسته ای با زمان گرفتن ورودی های در لایه اثراتی به دور از این می دهد و کار شبکه را

در یادگیری و همگرای سریعتر امکان می دهد به عبارت دیگر با استفاده از این کنشک توزیع داده های ورودی در

لایه با به روز رسانی وایزهای تغییر می کند و همان از یک توزیع زمانی به میانی و در زمانی که قابل

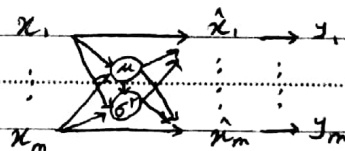
تنظیم است پس می تواند به بنای این شبکه به تغییرات توزیع داده های ورودی مقابله می کند. از دیگر مزایای این

کنشک می توان به هموار کردن بهینه سازی، امکان استفاده از نرخ های یادگیری بالاتر و در نتیجه همگرای سریعتر

دارا به بین اتی منظم سازی و کاهش خطر overfitting اشاره کرد.

$$\checkmark \frac{\partial L}{\partial \beta} = \sum_{i=1}^n \frac{\partial L}{\partial y_i} \frac{\partial y_i}{\partial \beta} = \sum_{i=1}^n \frac{\partial L}{\partial y_i} \quad (ب)$$

$$\checkmark \frac{\partial L}{\partial \delta} = \sum_{i=1}^n \frac{\partial L}{\partial y_i} \frac{\partial y_i}{\partial \delta} = \sum_{i=1}^n \frac{\partial L}{\partial y_i} \hat{x}_i$$



$$\left\{ \begin{aligned} \frac{\partial L}{\partial \hat{x}_i} &= \frac{\partial L}{\partial y_i} \frac{\partial y_i}{\partial \hat{x}_i} = \frac{\partial L}{\partial y_i} \delta \\ \frac{\partial L}{\partial \sigma^2} &= \sum_{i=1}^n \frac{\partial L}{\partial \hat{x}_i} \frac{\partial \hat{x}_i}{\partial \sigma^2} = \sum_{i=1}^n \frac{\partial L}{\partial \hat{x}_i} (x_i - \mu) \cdot \frac{-1}{\sigma^3} (\sigma^2 + \epsilon)^{-1/2} \\ \frac{\partial L}{\partial \mu} &= \sum_{i=1}^n \frac{\partial L}{\partial \hat{x}_i} \frac{\partial \hat{x}_i}{\partial \mu} + \frac{\partial L}{\partial \sigma^2} \frac{\partial \sigma^2}{\partial \mu} = \sum_{i=1}^n \frac{\partial L}{\partial \hat{x}_i} \frac{-1}{\sigma^3} (\sigma^2 + \epsilon)^{-1/2} + \frac{\partial L}{\partial \sigma^2} \frac{\sum_{i=1}^n -2(x_i - \mu)}{n} \end{aligned} \right.$$

$$\checkmark \frac{\partial L}{\partial x_i} = \frac{\partial L}{\partial \hat{x}_i} \frac{\partial \hat{x}_i}{\partial x_i} + \frac{\partial L}{\partial \sigma^2} \frac{\partial \sigma^2}{\partial x_i} + \frac{\partial L}{\partial \mu} \frac{\partial \mu}{\partial x_i} = \frac{\partial L}{\partial \hat{x}_i} (\sigma^2 + \epsilon)^{-1/2} + \frac{\partial L}{\partial \sigma^2} \frac{2(x_i - \mu)}{n} + \frac{\partial L}{\partial \mu} \frac{1}{n}$$

V

دوشنبه

Monday

27 Mar 2023

۱۴۴۴ (مهر) ۵

فرمانروای

$$X = \begin{bmatrix} x_1 \\ x_r \\ x_r \\ x_L \\ x_3 \end{bmatrix}_{\Sigma x_1}$$

$$W = \begin{bmatrix} w_1 \\ w_r \end{bmatrix}_{r \times 1}$$

(۱۵)

$$O = \text{conv}(X, W) = \begin{bmatrix} w_1 x_1 + w_r x_r \\ w_1 x_r + w_r x_r \\ w_1 x_r + w_r x_L \\ w_1 x_L + w_r x_3 \end{bmatrix}_{\Sigma x_1} = \begin{bmatrix} o_1 \\ o_r \\ o_r \\ o_L \end{bmatrix}_{\Sigma x_1}$$

$$\frac{\partial L}{\partial w_1} = \sum_{i=1}^f \frac{\partial L}{\partial o_i} \frac{\partial o_i}{\partial w_1} = \sum_{i=1}^f \frac{\partial L}{\partial o_i} x_i = \frac{\partial L}{\partial o_1} x_1 + \frac{\partial L}{\partial o_r} x_r + \frac{\partial L}{\partial o_r} x_r + \frac{\partial L}{\partial o_L} x_L$$

$$\frac{\partial L}{\partial w_r} = \sum_{i=1}^f \frac{\partial L}{\partial o_i} \frac{\partial o_i}{\partial w_r} = \sum_{i=1}^f \frac{\partial L}{\partial o_i} x_{i+1} = \frac{\partial L}{\partial o_1} x_r + \frac{\partial L}{\partial o_r} x_r + \frac{\partial L}{\partial o_r} x_L + \frac{\partial L}{\partial o_L} x_3$$

$$y = \begin{bmatrix} \frac{o_1 + o_r}{r} \\ \frac{o_r + o_L}{r} \end{bmatrix}_{r \times 1} \quad (1)$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial y_1} \sum_{i=1}^r \frac{\partial y_i}{\partial o_i} \frac{\partial o_i}{\partial w_1} + \frac{\partial L}{\partial y_r} \sum_{i=r}^f \frac{\partial y_i}{\partial o_i} \frac{\partial o_i}{\partial w_1} = \frac{1}{r} \left(\frac{\partial L}{\partial y_1} (x_1 + x_r) + \frac{\partial L}{\partial y_r} (x_r + x_L) \right)$$

$$\frac{\partial L}{\partial w_r} = \frac{\partial L}{\partial y_1} \sum_{i=1}^r \frac{\partial y_i}{\partial o_i} \frac{\partial o_i}{\partial w_r} + \frac{\partial L}{\partial y_r} \sum_{i=r}^f \frac{\partial y_i}{\partial o_i} \frac{\partial o_i}{\partial w_r} = \frac{1}{r} \left(\frac{\partial L}{\partial y_1} (x_r + x_r) + \frac{\partial L}{\partial y_r} (x_L + x_3) \right)$$

$$y = \begin{bmatrix} o_r \\ o_w \end{bmatrix}_{r \times 1} \quad (2)$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial y_1} \frac{\partial y_1}{\partial o_r} \frac{\partial o_r}{\partial w_1} + \frac{\partial L}{\partial y_r} \frac{\partial y_r}{\partial o_r} \frac{\partial o_r}{\partial w_1} = \frac{\partial L}{\partial y_1} x_r + \frac{\partial L}{\partial y_r} x_r$$

$$\frac{\partial L}{\partial w_r} = \frac{\partial L}{\partial y_1} \frac{\partial y_1}{\partial o_r} \frac{\partial o_r}{\partial w_r} + \frac{\partial L}{\partial y_r} \frac{\partial y_r}{\partial o_r} \frac{\partial o_r}{\partial w_r} = \frac{\partial L}{\partial y_1} x_r + \frac{\partial L}{\partial y_r} x_L$$