

I Subject: جایزه ۳

Year: Month: Day:

یادگیری درست روند معنی معنی چیست؟

فرآیند است این مدل با تعلم بر اساس های

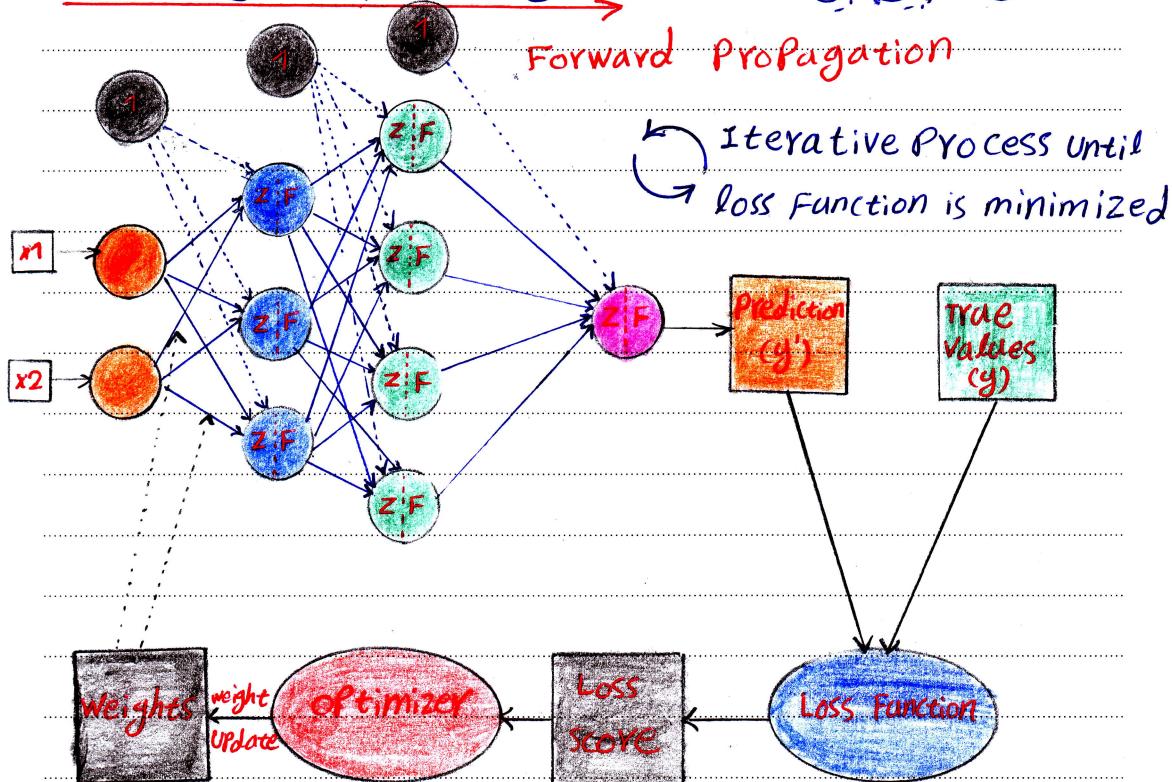
Loss

داخلی خود از مفهوم بروز رسانی عزت ها و بهینه سازی تابع خطا متعارضی.

تشخیص و پیش بینی الگوهای در داده های ورودی را بدست یافته اند

Forward Propagation

Iterative Process until
loss Function is minimized



Backward Propagation

Subject: / / Year: / Month: / Day: /

در حسابات قبل سبد های عصبی Forward می رویم به روی جلوه حرارت بین راهنمایی

اتفاقاً نهی افتاداین بوده ورودی های ۱ و ۰۲ وارد مدل مان کی سند و بالا میگارد

از خروجی مدل (\hat{y}) را می بینیم $\Sigma w_i x_i + b$

آنچهی در برداری مدل مانا میگیرد Backward نیز داریم

اتفاقاً نهی افتاداین است به روی محاسبه برای تراویح عوزن های آن بینتی کند و بعد

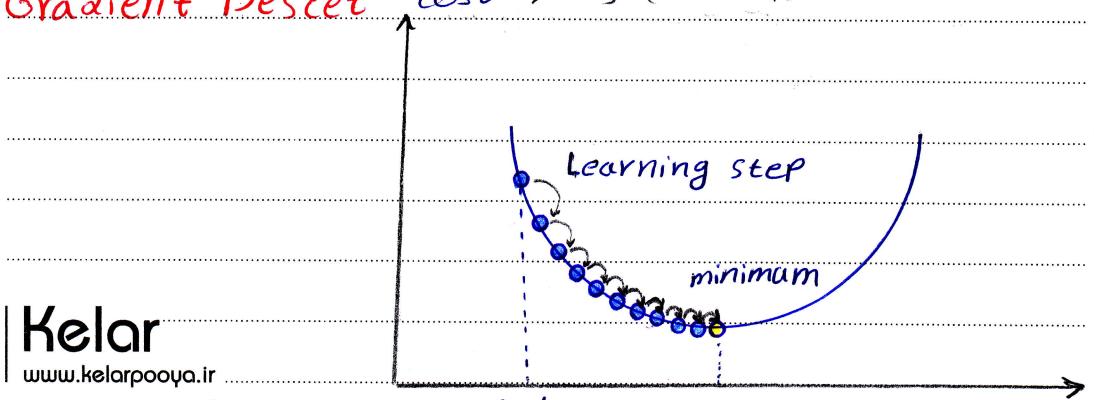
آنچهی در برداری مدل مانا میگیرد Backward و Forward

و loss فانکشن به کمین مقدار برداری میگیرد

optimization: این بینه بینی وزن ها و بارهای را بدست اوریم

و به الگوریتم این نام را به اینجا (Optimizer) میگیریم

Gradient Descent cost \rightarrow Loss (mse)



Kelar

www.kelarpooya.ir

Figure 4.4. the learning rate is too small

I Subject: ... Year: ... Month: ... Day: ...
 می باشد.

بی از الگوریتم کمایی به باعث چاهیش بسیار دن Loss مایی سود را

Descent می باشد.

هر این قدر در محور w وجود دارد ولی تعداد محدودی از آنها loss سان با این معنی ندارند.

اولین پارامتر θ (w) بصریت زنده است. به تبع آن مقادیر loss هم

بسیار سده و در نهاد رکرده هستند. سپس با تغییر θ کمان کار تکرار می شوند تا در نهاد نهاد نهاد

محض قبل برسیت آمدید است.

هدف کمان این است که بین تین پیشنهاد را نهاده کنند و این نهاده بعده برسیم.

از مرض لیم یکی را از لب، دلخواه، رحم، مژده، درزهای مختلف و باقای های دیگران

Learning step می نوییم بسته Min حرکت خواهد شد.

این باید "min" بسته min باید محدودی خوبی سریعی خواهد داشت

1 Subject: _____ Year: _____ Month: _____ Day: _____

:Hands on Machine Learning w/ Figure 4-3

in this depiction of gradient descent, the model

Parameters are initialized randomly and get
tweaked repeatedly to minimize the cost functions;

the learning step size is proportional to the slope
of the cost function, so the steps gradually get
smaller as the parameters approach the minimum.

داین تصویر از برآوران ترکیه پارامترها را در مسیر کمینه کردن دیده است

cost خطی برسانید. اگر cost خطی باشد، با هر چند مرتبه

cost با برآوران بازدهی نماید، پارامترها بحال پرداخته شوند.

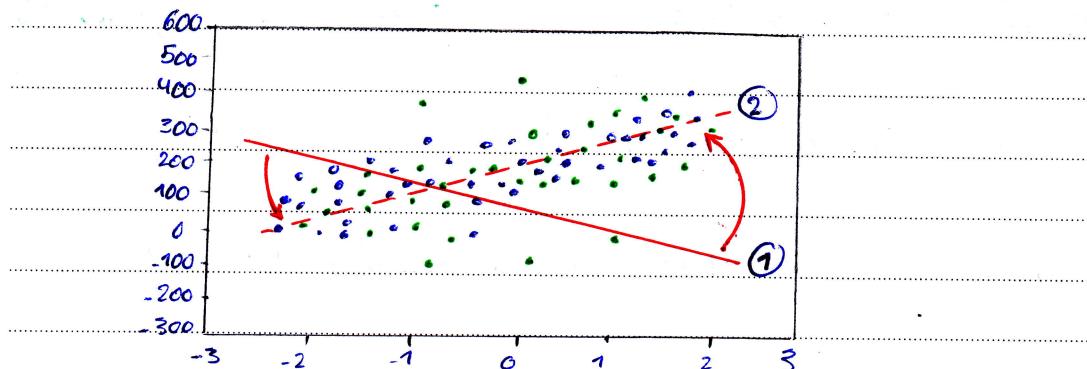
Learning Rate

Learning Rate برای هر چند مرتبه

Kelar

www.kelarpooja.ir

I Subject: _____ Year: _____ Month: _____ Day: _____



mean squared ERROR (mse) :

بهای داده ریسمانی خطا مان (MSE) بیان داده های داده شده است:

خط با تغیرات خالص ① و ② در اصل بارا متر کاری نتیجه ای انتخابی نیست.

فایلینگ (تحویل) خاصیت بهای داده های برای درست داده های داده شده است.

Leaner regression, Gradient Descent این اتفاقی است.

مقدار

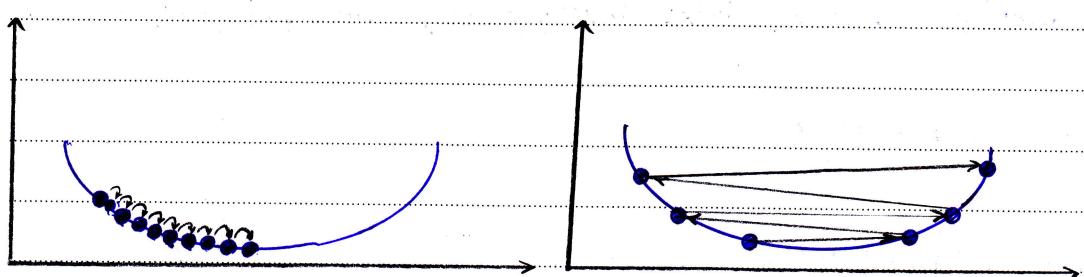


Figure 9-4. the learning rate is too small

Figure 9-5. the learning rate is too large.

I Subject: Year: Month: Day:

لریجیستری خانه خود برداری به $\hat{\theta}$ یا مینیم بررسیم learning rate

پس بازخواص این پاسخ را در انتخاب نتیجه از خارجی بزرگ انتخاب نموده از رو

نمیم تعداد بار زیارتی که پردازشی داشت و متوجه نمیم

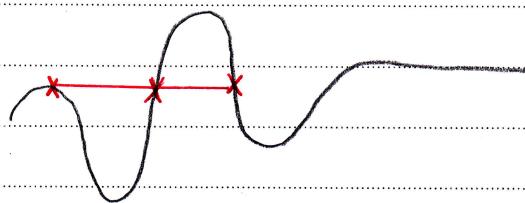
کوچکترین باندهن: رسیدن یا نزدیکی به \min این نورخ تردد کم ب معنی کوچکترین است converge

(خط) : زیان و ۲ تغییر مختلف بین نزدیک انتخاب تردد و یا عدم Non-concave و concave

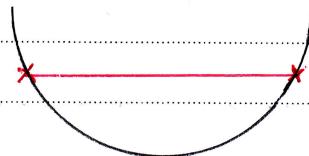
وصل شوند و این خط نزدیک را عقدا در ۲ تغییر قطعیت لست آنرا

این علاوه بر ۳ حق انتخابی نزدیک را جایی (برای) عدم قطعیت لست Non-concave

Non-concave:



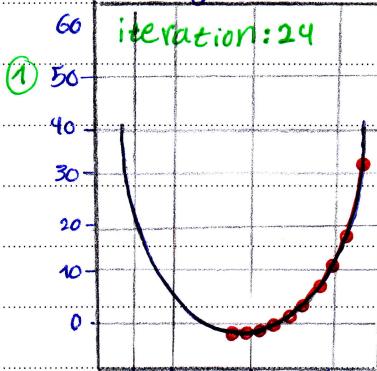
Concave:



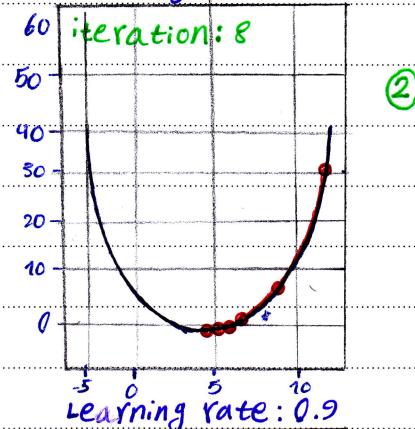
I Subject: Year: Month: Day:

*convex

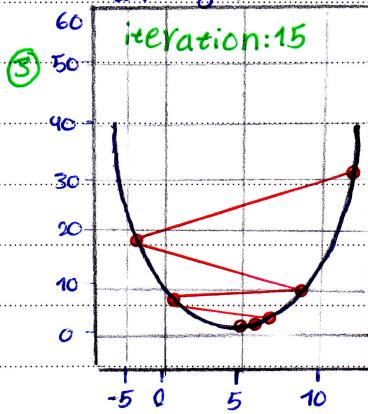
Learning rate: 0.1



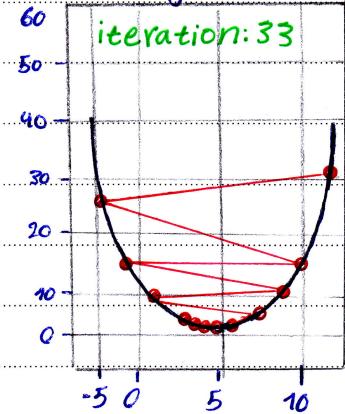
Learning rate: 0.3



Learning rate: 0.8



Learning rate: 0.9



۱. در سینه دادن یا میانگین کمای خیابانی را بخوبی انتخاب کنید و زمان زیارت حفظ کنید.

۲. تعداد نام کمای لصیری برداشت آنرا برسی کنید.

۳. به جای کمای از \min بروند و خود را را بخوبی درود بخواهید.

خود را را بخوبی درود بخواهید.

Kelar

www.kelarpooqa.ir

Subject: Year: Month: Day:

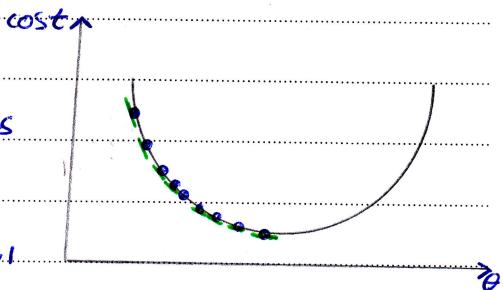
Gradient

گرادیان در ریاضیات و بحثه در زمینه بهینه سازی و داده لیری باشند مسأله دهنده

سب سازخ تغییرات باید تابع در پر نفعه خاص است و بعدها مسأله تراویان

برداری است مسأله باید چنون مقدار تابع در پر نفعه خاص تغییری لذت در چه جهت

لین تغییرات بین ترین مقادیر را دارد



که این قدر را مسأله دیگر سود سب خط در حال کم شدن

است تابعی در \min_0 بی سعد و در DL

اینجوی از خود گرادیان اسفلاین لندورن لایس محنت نهاده گرادیان کاهش پیدا

نمایه کنیم بآن گرادیان با همی لوین

$$\text{Gradient} \sim \nabla F(x,y) = \underbrace{\frac{\partial F}{\partial x}}_{x} \hat{x} + \underbrace{\frac{\partial F}{\partial y}}_{y} \hat{y}$$

مسفت تابع بسته ب \hat{x} مسفت تابع بسته ب \hat{y}

$$F(x,y) = x^2 + y^2 \quad \nabla F = 2x \hat{x} + 2y \hat{y}$$

I Subject: Year: Month: Day:

مسنون جزئی $\rightarrow f(x, y) = x^3 + y^2 \rightarrow \frac{\partial f}{\partial x} = 3x^2$

$$\rightarrow \frac{\partial f}{\partial y} = 2y$$

$$\left\{ \begin{array}{l} y = 8z^3 + z^2 + 5 \\ z = x^5 + 9 \end{array} \right.$$

مسنون زنجیره ای $\rightarrow \frac{dy}{dx} = ?$

آنکه مسند زنجیره ای باشد لازم بودی زنجیره ایست

$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial z} \cdot \frac{\partial z}{\partial x}$$

Gradient Descent (راحل)

① مقدار دفعی اولین پارامترها سے W کا رابطہ رندوم مقدار دھنیں۔

② کاسہ برادیاں سے بازنگاچ loss loss سے مسند بولیں۔

α : learning rate

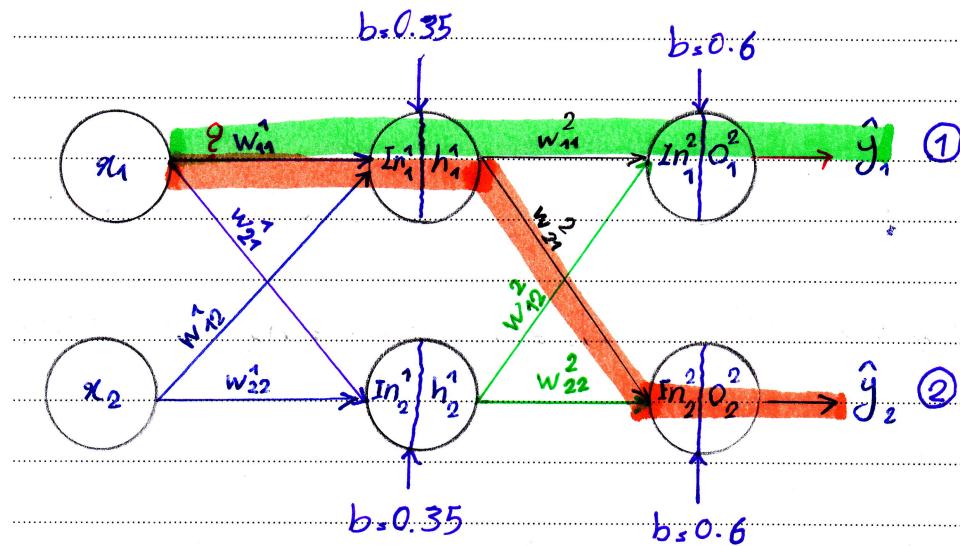
$$W_{\text{new}} = W_{\text{old}} - \alpha \frac{\partial L}{\partial W_{\text{old}}} \quad \text{ایک یعنی ہے} \quad ③$$

④ سڑھا جائے سے برآمد 2 و 3 راترسیدن (مسنون سدھ)

I Subject: Year: Month: Day:

حال برمیگردید، این نتایج را باز میکنیم.

Backpropagation in MLP:



$$W^1 = \begin{bmatrix} 0.15 & 0.2 \\ 0.25 & 0.3 \end{bmatrix} \quad W^2 = \begin{bmatrix} 0.4 & 0.45 \\ 0.5 & 0.55 \end{bmatrix}$$

$$b^1 = 0.35, b^2 = 0.6 \quad \hat{y}_1 = 0.01, \hat{y}_2 = 0.99$$

AF: F^1 : ReLU F^2 : Sigmoid $\eta_{h_1} = 0.05, \eta_{h_2} = 0.1$

? Update $w_{11}^1 \rightarrow$ SGD $\alpha = 0.01$

I Subject: Year: Month: Day:

① Feed Forward

② Loss Function

③ Back Propagation

Feed Forward ①

$$In_1^1 = W_{11}^1 x_1 + W_{12}^1 x_2 + b^1 = 0.3375 \rightarrow h_1^1 = \text{ReLU}(In_1^1)$$

$$= 0.3375$$

$$In_2^1 = W_{12}^1 x_1 + W_{22}^1 x_2 + b^1 = 0.5675 \rightarrow h_2^1 = \text{ReLU}(In_2^1)$$

$$= 0.5675$$

$$In_1^2 = W_{11}^2 h_1^1 + W_{12}^2 h_2^1 + b^2 = 1.1057 \rightarrow O_1^2 = \text{sigmoid}(In_1^2)$$

$$= \hat{g}_1 = 0.7513$$

$$In_2^2 = W_{21}^2 h_1^1 + W_{22}^2 h_2^1 + b^2 = 1.2297 \rightarrow O_2^2 = \text{sigmoid}(In_2^2)$$

$$= \hat{g}_2 = 0.7729$$

I Subject: Year: Month: Day:

$$\textcircled{2} \text{ loss function } \rightarrow \text{MSE} = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2$$

نحوه عرضی برای یک سوپر سامپل است که در SGD است.

\textcircled{3} Back Propagation:

برای یادگاری این روش را با ماتریس W_1 و بلوک های مجزا در نظر بگیر!

$$\frac{\partial L}{\partial W_1} = \textcircled{1} + \textcircled{2} = \left[\frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial O_1^2} \frac{\partial O_1^2}{\partial In_1^2} \frac{\partial In_1^2}{\partial h_1^1} \frac{\partial h_1^1}{\partial In_1^1} \frac{\partial In_1^1}{\partial W_1^1} \right]$$

+

$$\left[\frac{\partial L}{\partial \hat{y}_2} \frac{\partial \hat{y}_2}{\partial O_2^2} \frac{\partial O_2^2}{\partial In_2^2} \frac{\partial In_2^2}{\partial h_1^1} \frac{\partial h_1^1}{\partial In_1^1} \frac{\partial In_1^1}{\partial W_1^1} \right]$$

$$\frac{\partial L}{\partial W_1^1} = \left[\frac{\partial L}{\partial \hat{y}_1} \frac{\partial \hat{y}_1}{\partial O_1^2} \frac{\partial O_1^2}{\partial In_1^2} \frac{\partial In_1^2}{\partial h_1^1} + \frac{\partial L}{\partial \hat{y}_2} \frac{\partial \hat{y}_2}{\partial O_2^2} \frac{\partial O_2^2}{\partial In_2^2} \frac{\partial In_2^2}{\partial h_1^1} \right] \left[\frac{\partial h_1^1}{\partial In_1^1} \frac{\partial In_1^1}{\partial W_1^1} \right] =$$

-0.249 0.187 -0.217 0.174 0.3375
1 0.4 1 0.55 0.05

$$\ast (a \cdot b + b \cdot c) = b(a+c)$$

0.0006647

I Subject: Year: Month: Day:

حال برای مسنت بایریم:

$$\frac{\partial I_{n_1}^1}{\partial w_{11}^1} = \frac{\partial}{\partial w_{11}^1} (w_{11}^1 x_1 + \overset{0}{w_{12}^1} x_2 + b^1) = x_1 = 0.05$$

$$\frac{\partial h_1^1}{\partial I_{n_1}^1} = \frac{\partial}{\partial I_{n_1}^1} (\text{ReLU}(I_{n_1}^1)) = \begin{cases} 1 & I_{n_1}^1 > 0 \\ 0 & I_{n_1}^1 < 0 \\ \text{undefined} & I_{n_1}^1 = 0 \end{cases} = 0.3375$$

$$\left(\frac{\partial a_n^b}{\partial x} = abx^{b-1} \right) \text{ در } x=0: \text{ سود:} \begin{cases} 1 & x > 0 \\ 0 & x < 0 \end{cases} : 1*$$

در اینجا در $x=0$ سود نیز 0 است (حال برای مسنت بایریم) از همان‌جا در اینجا در $x=0$ سود نیز 0 است (حال برای مسنت بایریم) 2*

وجود ندارد.

$$\text{و مسنت: } 1 \leftarrow x \text{ در } \text{ReLU}(x) = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases} : 3*$$

ReLU در $x=0$ درست و در $x \neq 0$ بهتر است (زیانی در $x=0$ درست).

مسنت در $x=0$ نیز در اینجا بحث نمایش نمی‌شود.

خواهد شد.

$$\frac{\partial I_{n_2}^2}{\partial h_1^1} = \frac{\partial}{\partial h_1^1} (w_{21}^2 h_1^1 + \overset{0}{w_{22}^2} h_2^1 + b^2) = w_{21}^2 = 0.55$$

I Subject: Year: Month: Day:

$$\frac{\partial O_2^2}{\partial In_2^2} = \frac{\partial}{\partial In_2^2} (\text{sigmoid}(In_2^2)) = \sigma'(In_2^2) \times (1 - \sigma'(In_2^2)) = 0.174$$

$$(\sigma(x))' = \sigma(x) \times (1 - \sigma(x)) : 4 *$$

$$\frac{\partial g_2^1}{\partial O_2^2} = 1$$

$$\frac{\partial L}{\partial g_2^1} = \frac{\partial}{\partial g_2^1} \left(\frac{1}{m} \sum (y - \hat{y})^2 \right) = \frac{-2}{m} \sum (y - \hat{y}) = -0.217$$

$$\frac{\partial In_1^2}{\partial h_1^1} = \frac{\partial}{\partial h_1^1} (w_{11}^2 h_1^1 + w_{12}^2 h_2^1 + b^2) = w_{11}^2 = 0.4$$

$$\frac{\partial O_1^2}{\partial In_1^2} = \frac{\partial}{\partial In_1^2} (\text{sigmoid}(In_1^2)) = \sigma'(In_1^2) \times (1 - \sigma'(In_1^2)) = 0.187$$

$$\frac{\partial \hat{y}_1}{\partial O_1^2} = 1$$

$$\frac{\partial L}{\partial \hat{y}_1} = \frac{\partial}{\partial \hat{y}_1} \left(\frac{1}{m} \sum (y - \hat{y})^2 \right) = \frac{-2}{m} \sum (y - \hat{y}) = -0.249$$

يتبع بعدها 2 سود خود تابع ضرب المسنون : 5 *

$$(F^2)' = F \times F'$$

$$W_{\text{new}} = W_{\text{old}} - \alpha \frac{\partial L}{\partial W_{\text{old}}}$$

$$\Rightarrow W_{11 \text{ new}} = 0.15 - (0.1 \times (0.0006647)) = 0.15006647$$

1 Subject: _____ Year: _____ Month: _____ Day: _____

انواع Gradient Descent

1) **EPOCH**: یک دوره زمانی است و در آن مقدار مجموعه داده را باز می‌گذرد.

بررسی مدل مجموعه داره های معزوف می‌بینیم.

2) **Batch**: یک یا چند مجموعه ای از نمونه های داده است که در یک مرحله از الگوریتم درآمدان دست داده می‌شوند.

* بخطابی اینه مدل داده های را استناد نمی‌نمایم بلکه داده های دنباله داری داشته که این 8 تایی تضمین شده.

در مرحله (Epoch) یک اراین دسته های را می‌گذرانند مدل می‌سازند اینها بعده می‌گذرند.

دسته داده "Batch" نویش.

سطل دوستان

3) یا در آن مرحله W_{new} چه نفعی داشت؟ (درآمدانی مدل)

اگر در کامپیوتر مدل لفته سه ۷۰۰۰ مقدار پردازی بود از W_{old} مقدار بیشتری

زیاد یافته شد. اگر مثلاً ۰.۱ بجا ۰.۹ باشد ۱۰۰۰ بعده مقدار ۰.۶۶ بجا

$W_{new} = 0.1 + 0.9 \times 0.1 = 0.00066$

I Subject:

Year: Month: Day:

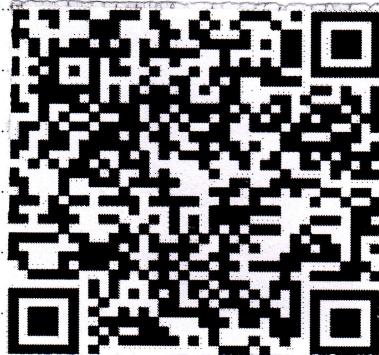
QR دیست دارم و هر کدام یک سپر دارند. با مراجعه به خود رسیده ام برای این دسته از ماده.

زیرا صفحه لیت ها ب این درجه وصل شده و توانید دلو تو خیفات استار + تو خیفات

لطفاً این را بازسازی کنید جی تی GPT رفته سد دانلور لش.

<https://github.com/Mohammad-Ali-Malekzadeh/>

Deep-Learning-Filoger



Gradient Descent

Stochastic Gradient Descent (SGD)

در این روش هر ادیار یک کوچک نداده به صورت جداگانه یا میتوان

Kelar
www.kelarpooja.ir

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

I Subject: _____ Year: _____ Month: _____ Day: _____

Batch Gradient Descent

در این روش برای این مجموعه داده برای تعریف مماسه بی سود

mini-batch Gradient Descent

در این روش ترکیب از دور و نیم قبل است که این صورت برای این برآمدگان برداشت کند

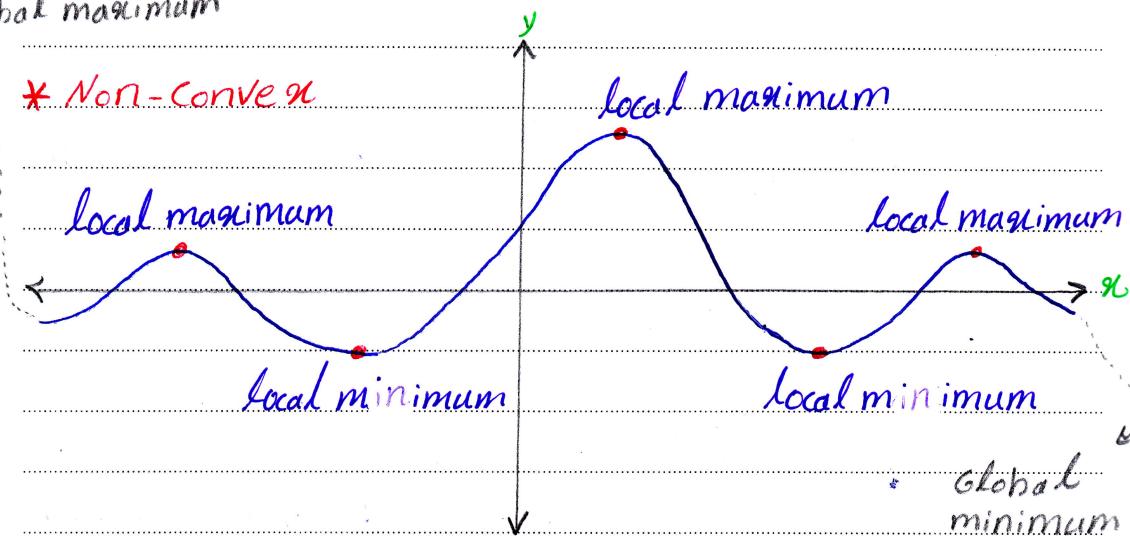
تجزیی از داده ها (مینی بچ ها) مماسه بی سود

معiar مقاييسه	Bath Gradient Descent	stochastic Gradient Descent	mini-Batch Gradient Descent
ماسه بی تراویان	نمایماده ها	خرنخه بعد سجد اگان	دسته های کوپ از داده ها
سرعت	کند	سریع	تعادلی بین سرعت و دقت
نیاز به حافظه	زیار	کم	متراز Batch و سین متراز SGP
دققت تراویان	بالا	نویز بالا	تعادلی بین نویز و دقت
نوسانات درجهت	کم	زیار	متراز SGP و سین متراز
فرادیان	نوسانات زیادی تواند از	نسل است به mini کا	تعادل بین ممکن تری و نویزها
کمتر ای	محالی بر سر	مینیمیم های علی جلو لیری لان	پیچیده تراز ۳ درس دیگر
پارهه سازی	ساده	ساده	نامناسب داده های کوپ
کاربرد	نامناسب برآورد داده ها و مسائل	نامناسب داده های سیز بزرگ	SGP و Bath بین
زن برع	زیار	کم	

Subject: Year: Month: Day:

global maximum

* Non-Convex



اگرچه در جهادی دستیابی به دست آمده باشیم، اما اینها local max و min هایی هستند.

اما معمول (global) (که بین دو مترین دستوراتی مقدار را داشته باشد).

در پایان اوضاع local minimum یا local maximum را دریافت.

برخورد با لغایل ها سچالش می خورد و در آن لیزی لذت اصلی خود تواند تشخیص

کردن باشیم. در نهایت لغایل است و من توانند آزاد آن ببرون. ایده بینی طوبایی دفعه هشت

I Subject: _____ Year: _____ Month: _____ Day: _____

Gradient Descent خالص دستور

:Local Minimum

خالص الورست میان است در نتایج این اکلی لیست بین تین نفع

دنباع هزینه نیست.

:Global Minimum

خالص تابع هزینه بین تین مقدار خود را رسید و همچو نایي اورست است

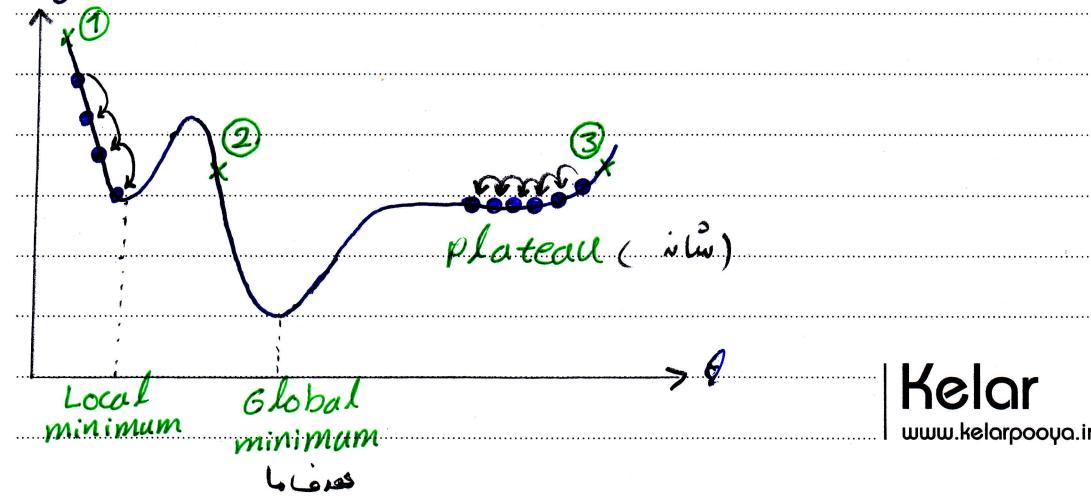
:Plateau

ناجیانه در این تابع هزینه تغیرات کم در ورودی اورست میان است به نظر جوت

*non-convex

لزیال بر بینند.

Cost



1 Subject: Year: Month: Day:

نمکل / چالش

توضیح

راحل ها

برآوردن در مقادیر اولی استفاده از الگوریتم های بهینه سازی می باشد. برآوردن بر اساس $\nabla f(x)$ است. برآوردن بر اساس $\nabla f(x)$ برآوردی علی بجای نقله ای است. Momentum نیز همین روش است که در آن β بهینه شده است.

Nestrov Accelerated Gradient

با استفاده از شبکه های عصبی عمیق با ساختار های مختلف.

انتخاب فرخ یادگیری ایجاد حالت استفاده از رویس (زرویس) های تطبیقی است. بیشتر از آن دو گامی سرعت فرخ یادگیری می باشد. Adam و AdaGrad، RMSprop و گرادیان فرخ یادگیری بعمور دینامیکی می باشد.

محنت تعلم یادگیرها برای مسائل تهدیدی یادگیری پردازه استفاده از زرویس های بهینه سازی تتفاوت فرخ یادگیری و سایر یادگیرها خود را تقویت می کند. hyperparameter می باشد. می باشد. می باشد.

با استفاده از زرویس های بهینه سازی

Grid search

Random Search

مقایسه نظری برآوردهایی برای ایجاد استفاده از زرویس های داده های اینجا از مجموعه داده های در هر مرحله Mini-Batch می باشد. می باشد. می باشد. می باشد.

Kelar

www.kelarpoya.ir

I Subject: Year: Month: Day:

ابتداً برای این سورس
واسطه از تکنیک های تغذیه سند
برای پردازش مولازی درجه.

حساس به مقدار اولیه
نتیجه نمای مدل است بسته استفاده از درس های بینه سازی
به مقدار اولیه وابسته باشد
با اولویت برای تنهایی و استثنایی
نمونه در شبکه های عصبی
لتریب هدایت اولیه مانند استفاده
از توابع فعال سازی و پس پردازش
مناسب

نقاط مستحبی از نظر در بخش مدل است
در بخش مدل از درس های بینه سازی
برای این سوال کمینه باید خاص باشد
Batch Dropout به ترتیب باعث تغییر مدل می شود
پس از درس های بینه سازی و استفاده
از درس های پیشرفته تر برای
پس پردازش دارد.

حال چگونه توانیم بین چالش های خود را در آنها ابراز کرد؟

مشکل - مسائل که برای حل این مسائل

Scaling *

Learning Rate) *

I Subject: Year: Month: Day:

* استفاده از روش های پیشرفته‌تری با AdaGrad و RMSProp و Adam می‌باشد.

مُومنِتوم (Momentum) *

ابدا مدار دادن وزن‌ها در نقاط مختلف (Random Initialization) *

دینونه از Scaling را دیدم (Normalization) است

لیکن Learning Rate می‌توانم local می‌توانم متناسب با سایر کم از دام های

خلاص نمایم و میریم از آن خارج نموده

روش ایترین است که واحدی به اسم مُومنِتوم به معنی حالت در آن پایه ایس

موابد نمایم افنا فریم

آن تغییر با جریت بلویم که مقادیری W خیلی تواند در بین افرازین یا نافرازین در

قطعی داشته باشد. به مرور آنچه نیز قبل نماید لذت... سعادت کمای می‌توانستم

متوازن باشد در نظر بررسی Random Initialization

I Subject:

Year: _____ Month: _____ Day: _____

که زمان از زندگی cost ماندن داشت

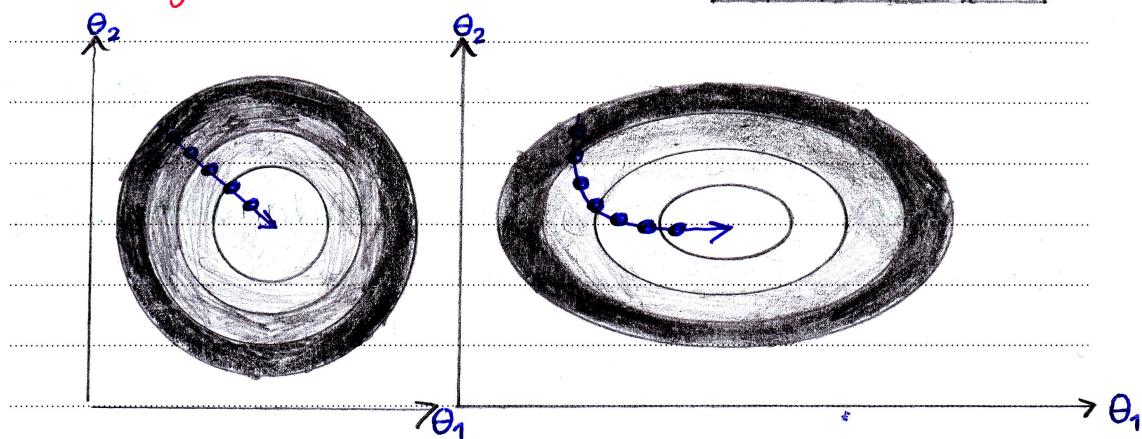
Scaling \rightarrow Normalization

Figure 9-7. Gradient Descent with (left) and without (right) Feature scaling.

در خود را بخواست راستی مانند بالای تصویر نداشتم و در پیشی راستی زیادی نداشتم.

پس از برخورد با تابع هدایتی، این کار را بخواهیم کرد که $J(\theta)$ (دراخانه) این

بردن را در بین مجموعه ای از θ ها کم کرده باشد $\Leftrightarrow \theta$ ها بهم در بین رینج قرار گیرند (محدوده استمر

$J(\theta)$ کمی باشد