



# Homework 3

Deep learning  
Dr. Mohammadi

## مشخصات

رحمت الله انصاری	نام و نام خانوادگی
404722028	شماره دانشجویی
کارشناسی ارشد هوش مصنوعی دانشگاه علم و صنعت ایران	رشته و مقطع و دانشگاه
<a href="mailto:rahmat.ansari.dev@gmail.com">rahmat.ansari.dev@gmail.com</a>	ایمیل

## ساختار فایل‌های پاسخ

- فایل answers، که همین فایل است، شامل
  - مشخصات دانشجو
  - پاسخ‌های تشریحی به ترتیب سوالات
  - گزارش notebook های experiment ها
  - رفرنس‌ها
- Practical
  - فایل notebook1-hw3-students.ipynb
  - فایل notebook2-hw3-students.ipynb
- پوشه helper (شامل فایل‌های کمکی، در این تمرین شامل کد مربوط به سوال ۲ بخش تشریحی)

(loss) لَا يُنْهَا بِحَرَقَةٍ إِنَّمَا يَأْتِي لَكُم مِّنْ أَنفُسِكُمْ ۚ ۖ

weights do not generally increase along with just older

mathematical problems involving Bases 3

لارج این یا لارج این دیگر نهاده و مجهز یا لارج این دیگر

وَمِنْ أَعْلَمِ الْأَعْلَمَ بِهِ مُحَمَّدٌ

وَمِنْهُمْ مَنْ يَعْمَلُ مُحْرَماً فَلَا يَرْجِعُونَ (٢) -

is Gay Bank Drug and Jewelers the one with the glass?

g (Digital dot = 8) and first syllable will be read.

Envolgimento perifilar em S. (cistoscelidae)

لذکریں بے انتہا جیسا کوئی مسازی یا کوئی تحریر تحریر اخراجی دفعہ انداز  
global minimum SGD کو سمجھو گے۔ (کمترین دفعہ)

with the

- contour plot  $\hat{S}$  vs  $y$  and  $z$  - do you notice anything?

الله اعلم بالمراد بـ بـ بـ بـ بـ بـ بـ بـ بـ

وهو ينبع من مفهوم المعرفة المترافق مع المعرفة المترافق

الآن يجري التحقيق في هذه المجموعة من الأحداث، وبيانها في جلسة مجلس الشورى يوم 30 مارس.

~~140-3670~~

• Top-down (Top-down) طريق العودة من الأعلى

تقریب طبقه و مکان روی گان =  
 این داده را که نویسندگان در آنها تعریف کرده اند برای درست شدن  
 گنجایش دارند. در اینجا این داده را با loss function می نویسند.  
 (MSE) درست نمودن داده ها را در اینجا درست نمودن داده ها  
 تغییر نمی کنند.

and minimum value of loss function که تابعی است که بازگشت

۱-۲) تابع این داده ها در داده های ورودی و loss function بعنوان  
 یک تابع که این داده های ورودی را بازگشت می کند. این تابع  
 این داده های ورودی را با تغییر داده های ورودی و loss function  
 تغییر نمی کند.

If we want to train a loss function in the fitting (دستیابی)

آنچه ایم اینکه loss function را در داده های training و test می کنیم  
 sharp minimum را بجستجو کنیم اما این overfitting  
 خواهد داشت. این داده های training را در داده های test خود  
 بازگشتی داشته باشیم (دستیابی داده های test) این داده های  
 test را در loss function اضافه کنیم.

این optimizer را loss function optimizer می نویسند.

این optimizer را loss function optimizer می نویسند. این داده های  
 training را در loss function اضافه کنیم (generalization) این داده های

بهینه سازی می کنیم.

and not

:= 18, 200

لذلك يكتب  $\lambda$  في  $L_2$  لـ regularization function.

پیر کیمیہ

مکاریں کو نہ سمجھنے والے بھی ٹیکنالوگیز اور وائیبکس پر

مکتبہ احمدیہ مسیحیہ

After I had a short walk I got tired & early stopping.

جاءت هذه الدراسة لبيان مدى دقة وفعالية التقييمات المقترنة (التحقق والتحقق المعاين) في تقييم مهارات القراءة لدى طلاب المدارس.

Final) Tools for data augmentation

وَالْمُؤْمِنُونَ لَهُمْ نُورٌ فِي الظُّلُمَاتِ وَلَهُمْ أَنْوَافٌ

flipping up, I thought a learning rate sheet -

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

4/2

learning rate = 0.1

$$b = +1 \quad \Rightarrow \text{and also} \quad$$

$$\text{beta} = 0.9$$

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

batch size = 2

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

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

(stochastic gradient descent) 2nd

فرموده و میانگین

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

first Batch

اولین

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

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

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

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

گرادیان و سلسله مراتب

$$\frac{2}{n} \sum e_i \times \frac{\partial \hat{y}}{\partial w}$$

این کسی این چیز را پیدا نمی کند

مشتق صفتی = مشتق

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial w}$$

مشتق خطا نسبت

نسبت وزن

$$\nabla w_1 = \frac{1}{2} (2(-8)(1)^2 + 2(-8)(2)^2) = -8 \cdot 32 = -256$$

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

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

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

مقدار و خوبی

$$w_1 = 1 - 0.1(-256) = 1 + 25.6 = 26.6$$

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

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

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

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

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

⇒ Batch  $\Leftarrow 2$  نمونه

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

$$e_1 = 1.8 - 11 = -9.2$$

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

$$e_2 = 9.2 - 11 = -1.8$$

⇒ داده های آموزشی

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

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

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

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

⇒ داده های آزمایشی

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

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

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

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

SGD + Momentum  $\Leftarrow$

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

سینه SGD در سیستم های آموزشی

$$\nabla w = (-40, -8, 8, -16)$$

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

(epoch 1) 1st iteration

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

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

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

(epoch 2) 2nd iteration

$$\hat{y}_3 = 1.4(0)^2 - 0.92(2)^2 - 1.8(0)(2) + 1.16 = -2.52$$

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

$$\hat{y}_4 = 1.4(-2)^2 - 0.92(1)^2 - 1.8(-1)(1) + 1.16 = 2.72$$

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

(epoch 3) 3rd iteration

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

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

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

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

(-4, -0.8, 0.8, -1.6)  $\in V_1$ , next place in  $V_2$ , etc

$$v_{w_1} = 0.9(-4) + 0.1(-1.28) = -3.728$$

$$v_{w_2} = 0.9(-0.8) + 0.1(-55.36) = -6.256$$

$$v_{w_3} = 0.9(0.8) + 0.1(1.28) = 0.848$$

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

(epoch 3) 3rd iteration

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

$$w_0 = -0.92 - 0.1(-6.256) = -0.2944$$

$$w_1 = -1.08 - 0.1(0.898) = -1.1648$$

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

۲- (الف) وقت نزف پادین (Over-shoot) برابر با مقدار اضافی که در مراحل اولیه فرآیند ایجاد شده است.

(Over-shoot) از نوع دینامیک (overshoot time) برابر با مدت زمانی که مقدار اضافی که ایجاد شده است را برای این فرآیند بود.

۳- (ب) وقت افتراق (Divergence) برابر با مدت زمانی که مقدار اضافی که ایجاد شده است از مقدار اصلی خود جدا شود.

۴- (ج) تابع خطا (Loss function) میان طرزهای مختلف ایجاد شده است.

۵- (د) وقت نزف پادین (Over-shoot time) برابر با مدت زمانی که مقدار اضافی که ایجاد شده است را برای این فرآیند بود.

۶- (ه) تابع خطا (Loss function) میان طرزهای مختلف ایجاد شده است.

۷- (ج) مقدار اضافی که ایجاد شده است در هر یکی از طرزهای مختلف ایجاد شده است.

۸- (د) مقدار اضافی که ایجاد شده است در هر یکی از طرزهای مختلف ایجاد شده است.

(saddle)

۳- ۲- نکات زیاده تر در فناوری AI میگذرد که در این بحث امتحان شده است که

عبارت مزبور (saddle point) این امر نتایجی است و بعدها مطالعه است و بعدها مطالعه

در مقاله گذشتند روشی دارد (saddle point)

در درجهات سینه او به بایرن دارد (ماکسیمم)

نکل آن هم شبیه زیست است

: SGD رختر

الگوریتم SGD آنرا برای این مسأله کلیاً لطفاً در کناره میگذارد.

زیرا گرادیان (gradient) نتیجه بحث است در نتیجه سرعت آپسی وزن

تحویل این معنی دارد. این معنی که چون یک داده را در یک بخش

نویزه آلته جویی SGD روی داده های کوچک (Batch) انجام می شود

نویزه های از داده های ممکن است تا این پاره های داده های معمولی نباشند

که باعث خوبی نیست و نهایت

برای آنکه adam را در مورد adam، نسباریستون دارد

درینکل قبل از آن معرفی شد. adam = momentum

نمیتواند مقدار باعث می شود بلکه در نظر ممکن است تا وقتی که

آن عبور کند

آنکه adam = Adaptive learning rate

adam درینکل که گرادیان خوبی جهای بسیار کوچک است

باید میگردید که این داده های کوچک در آن مهندسی

و بزرگ ترین که بتواند سریع تر از adam نباشد. در نتیجه

در adam از نقاط زیادی بسیار سریع و موفق تر از SGD عمل می کند

لکھتے ہیں (smooth) اور دوسرے لکھتے ہیں (noise).

Or Sing. Cylindrical - and Batch gradient  $\rightarrow$  very smooth flow  
Descent

وَالْمُؤْمِنُونَ إِذْ يَرَوْنَهُمْ يَقُولُونَ إِنَّمَا نَعْلَمُ مَا بِأَعْيُنِنَا

کتابوں ایک سلسلہ کی وجہ کاریوں و اقتضائی و دعیفے کل ایک پوچھ دادا رہتے ہیں

وَلِكُلِّ مُؤْمِنٍ دُرْجَاتٌ مُّتَّقِيَّةٌ

Energy (losses) mini-Batch gradient Descent, Vegas. Many steps -

د. سید علی‌محمد حسینی (متولد ۱۳۲۷) از اولین افرادی است که در این زمینه فعالیت می‌نماید.

ای زیر مجموعه کوچکترین از گروه و گروهی است و در این

لے کر اپنے بھائی کو سمجھا۔ اس کے بعد اپنے بھائی کو دیکھنے کا کام پورا کر کر اپنے بھائی کو سمجھا۔

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

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

بیان ۷ میلی کوچک:

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

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

وَالْمُؤْمِنُونَ إِنَّمَا يَرْجُونَ حِلَالاً

$$J(w^{t+1}) \approx J(w^t) + \underbrace{(\nabla J(w^t))^T}_{\text{جواب احتمالی}} \cdot \underbrace{(w^{t+1} - w^t)}_{\Delta w}$$

در نظر مطلع

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

جوابگاری 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 تری سمت راست را در حالت متوسط نماییم

$$\nabla^2 V = \| \nabla V \|_2^2 \text{، کوئی دو چیزی نیست}$$

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

نیز باید

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

$$\underbrace{\begin{array}{c} \text{نمایش} \\ \text{کوئی} \\ \text{دوسان} \\ \text{نمایش} \\ \text{کوئی} \\ \text{دوسان} \\ \eta > 0 \\ \text{داریم} \end{array}}_{\text{نیز باید این مقدار را کوئی داشت}} \quad \underbrace{\begin{array}{c} \text{نمایش} \\ \text{کوئی} \\ \text{دوسان} \\ \text{نمایش} \\ \text{کوئی} \\ \text{دوسان} \end{array}}$$

نیز باید این مقدار را کوئی داشت

باید مالات را کوئی داشت

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

proof 7/1-5

SGD (stochastic gradient descent) کاری کوئی توجه نهاده ایم Adam ویرایش

sparse یعنی کوئی کوئی داده نداشته باشد. این کوئی کوئی داده توپیکی باشد

که داده هایی که این داده را تشکیل می دهند از داده هایی که این داده را تشکیل می دهند

کوئی کوئی داده توپیکی باشد. این کوئی کوئی داده توپیکی باشد

## ~~Physical Cognitons~~ P<sub>T</sub>

Exponential Moving Average

$g_\theta \leftarrow \nabla_\theta f_t(\theta_{t-1})$  جواب مدرس (ز)

$$m_f = \beta_1 \cdot m_{f-1} + (1-\beta_1) \cdot g_f$$

$$W_t \leftarrow \beta_2 \cdot W_{t-1} + (1 - \beta_2) \cdot g_t^2 \text{ (مربع الخطأ - متوسط الخطأ)}.$$

$$\hat{m}_g \leftarrow m_g / (1 - \beta_1^t) \quad \text{جایگزینی مولکولی}$$

$$\hat{w}_j \leftarrow w_j / (1 - \beta_2^t) \quad \text{ranging from } 12 \text{ to } n$$

$$\theta_t \leftarrow \theta_{t-1} - \alpha \cdot \hat{m}_\theta / (\sqrt{\hat{v}_\theta} + \epsilon) \quad \text{Durchschnittswert}$$

(Confidential information contained herein is proprietary to [REDACTED])

جیلگیر مولانا ۲۰۰۷ ۱۲-۵

(٦) **العنوان** ينبع من الكلمة الأولى في العنوان.

$\text{Gr}^{(0)}_{\mu\nu}(w_0)$  ( $w_0$  یک پوزیشن است) ( $m_0 = 0$  و  $v_0 = 0$ )

Contra (in eos) sibi B<sub>2</sub> & B<sub>1</sub> ad ipsos Colijg. Emig. &c. valgt.

most interesting

وَالْجُنُوبَ وَالْمَغَارَاتِ وَالْمَدَائِنِ وَالْمَدِينَاتِ

مکان ایسے ہے جو اپنے بھائیوں کے لئے اپنے بھائیوں کے لئے ایسا ہے۔

$\frac{1}{\sqrt{2}}(B_1 - B_2)$  Lijens, RMSProb 5.2, and Dr. Hergenreider

وَهُنَّ أَقْرَبُ إِلَيْنَا مِمَّا يَرَوْنَ

✓ 2014 R.M.S.P. vs. Christopher Adams (C)

وَجَعَلَ لِكُلِّ أَنْوَاعِ الْمُجْرَمِينَ إِلَيْهِ مُنْسَكًا

Figures can also be made of RMSProp to our liking, as shown.

Der 10. Februar 1918 hat mir der Herr B. S.

(b) Reinstatement and reinstatement 18-5

std + moment ( )

لهم افتح دروازه سلام و انتقام مني

• *Nérot Mometronidazole*

g = 1 - \frac{1}{2} \left( \frac{\partial \ln f}{\partial x} \right)\_0^1 \text{ by formula of total differentiation}

right angle. I can't say it's a right angle because it's not.

Do you want to continue?

Adagrad ( $\rightarrow$ )

ال Sparse gradient descent افضل adam

(Vg) . امثلة على Adagrad

الخوارزمي (  $\sum g^2$  ) افضل adam

adam و دردشة سوقة

الخوارزمي (  $\sum g^2$  ) افضل adam

الخوارزمي (  $\sum g^2$  ) افضل adam

RMSprop ( $\ominus$ )

الخوارزمي (  $\sqrt{g^2 + \epsilon}$  ) افضل adam

Iterations 6 (  $\frac{\epsilon}{\sqrt{3-5}}$  )

الخوارزمي (  $\sqrt{g^2 + \epsilon}$  ) افضل adam

CNN C-SPAN is owned by Adam & Jerry P. Levin & the <sup>Abe F. Peabody</sup> (G)

وَهُوَ مُؤْمِنٌ بِأَنَّهُ آدَمَ مُرْسَلٌ مِّنْ أَنْفُسِ الْأَنْوَارِ  
— وَهُوَ CNA، وَهُوَ آدَمَGrad، وَهُوَ إِلَيْكُمْ لِي

and get it to you.

Options 25-5

نہ آورج و دستورد : بحثیت روئی ملکی بھین ساری تھا صحنِ سربراہ اول

بـ اـ لـ اـ سـ تـ حـ يـ وـ هـ اـ رـ تـ حـ يـ وـ هـ اـ رـ تـ حـ يـ وـ هـ اـ رـ

## • Medical Record

حکم ایجاد کردن مکانیزم: این کار در برخادر باشد

وَكُلُّ مَا يَرُونَ لِمَنْ يَرِي  
وَلِمَنْ لَا يَرِي

Program) Automatic Annealing Signal (Wiggle)

(inf. c. 18-20) शोक

فاطمة نصر وصلح

(intuitively interpreted) as the rules of inference.

(invariant to rescaling)  $\mathcal{D}_{\text{G}} \leftarrow \mathcal{D}_{\text{G}} \cup \mathcal{D}_{\text{G}}^{\text{new}}$ .

کارخانه هایی که در اینجا تولید می کنند (میراث بازیافتی) از جمله

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

وَالْمُؤْمِنُونَ الْمُؤْمِنَاتُ الْمُؤْمِنَاتُ الْمُؤْمِنَاتُ الْمُؤْمِنَاتُ الْمُؤْمِنَاتُ الْمُؤْمِنَاتُ

(Non Convex)  $\Rightarrow$   $\text{non-convex} \Rightarrow \text{No Global Optimal (Convex)}$

## نکالیں گا تو جو کہ

أبا إسماعيل زاده حنفية

کارہ ایسے کہ مفہومی سلسلہ لڑک و بارہاد بالا نہیں آئے وانہا

Full gradient by 40 Epochs

1/ Notcholle 1 loss

دراج آنچه با هر یک batch از لایه مختلف

درست پیش از آنچه با هر یک batch از لایه مختلف درست شود

به داده های 21, 25, 120, 250, 1000

= 10 Loss =

(136.65 درست پیش از هر یک batch = Full batch (100) -

(172.39 درست پیش از هر یک mini batch (1) = 10 درست پیش از هر یک

روند پیش از هر یک داده 21, 25, 120, 250, 1000 درست پیش از هر یک داده 172.4 + 172.9 = 345.3

2 Full batch پیش از هر یک داده = 2

? پسی epoch کم = درست 100 وزن 40 درهم

آنچه پیش از هر یک داده 100 درست

فروخته شود (30 درست از هر یک داده 100 درست)

باشد

آنچه احتساب در نتایج نهیج و کارهای دقیق (بیوگرافی) در

پیش از هر یک داده 2 (Scalp Points) یعنی 100 درست

از هر یک داده

mini batch (1) 5, 7, 10, 15, 3

- ضرکاریون بیانیت (Regularization effect) دوستی داشت و از پرداخته باز رسانی کرد  
و نوع تاثیر هر سیار زیاد است که ممکن است در آنچه در آنکه برای این تاثیر از این اثر  
کوچک باشد ایجاد شدید در تحقیق کار این اثراست این نظر

و همچنانچه sharp minimum (stochasticity) در گذشت (generalization) یا عدم گذشت (overfitting) یا عدم گذشت (underfitting)

در یادداشت (1.3 t2) mini-batch gradient descent نتیجه ای این است

که اگر این اثراست را در آنکه ای از آنها را از این نظر در نظر گیری کنیم

برای مثال در full batch gradient descent

Momentum (جذب) و SGD (experiment 2)

نمایش داده شد

نمایش داده شد (convergence speed) که این اثراست

نمایش داده شد

نمایش داده شد (SGD + Momentum)

(8 epochs) نمایش داده شد (SGD + Nesterov)

(8 epochs) نمایش داده شد (SGD (standard))

نمایش داده شد (SGD + Momentum) نمایش داده شد

نمایش داده شد (SGD + Momentum, learning rate) نمایش داده شد

نمایش داده شد (SGD, learning rate) نمایش داده شد

## تيلمذ رفتار گواهان ۔ ۲

مکانیزم در ماده مخصوصاً (Momentum Conservation) پلتراس (Plasterus)

طعن: (تعديل بـ من بعد دعوى ٨٧٨ رقم ٦١٢٠٣)

Momentum  $\rightarrow$  is also called rest or momentum -

(field) 11-12-1978 Nesting along coast of

Years (look ahead) suggest best nesting site

Question 1 - binary momentum L = 17.96

مودتیم که این روزها بسیار خوب است.

It was first organized by George Washington.

یافر پالر از سینمای فرانسه

Long Island City 5BD

## Costs & Income Experiments

سید علی

(~~Stony~~ Inorganic, Non-Magnetic) (LR = 0.01) is lost.

- 1 -

(% 24.36), symbiosis, *Rhizobium* 560

دینی درجی نظریہ nestor و momentum کے لئے

Chit & Chit, 2nd floor (145 sqm) - 12373

Writing Good English Sentences

سبت ۱۵۸ میلادی (560+Mammoth)

دیکنیسیتی ایکاؤنٹینگ : (LR 20.5) ڈیکنیسیت

أَنْتَ مَنْ تَرَى

SGD: همچنان روشی است و باید دقت (۱۴۰) صفحه کتابخانه اینستیتو

$\sin 10^\circ \approx 0.1736$  (2 R.S.O.)

- Island of Puglisi: Nestbox measurements

مکالمہ رفتار دینا یاد کرو۔

گروهای هم روزانه بزرگ سنتومر نزدیک: وقتی نزدیکی بزرگ سنتومر نزدیک است (0.5-1)

این پدیده را بیزگ بخوبی می‌دانیم و در اینجا نیز می‌توانیم آن را در اینجا در خود داشتیم.

استحکام SGD = کاریابی و کارگذاری استاندارد چون حافظه اندیز

دستفهای ملکه نیز نیز آن میگیرند و این دو دسته

دین و نیز مذکور شده اند و ممکن است در این دسته قرار گیرند.

میرخ بادگیری با احتمال بسیار کمتر از ۵٪ در میان هفتاد

هسته ای سیار سریع ترین دارند. تنظیم مقیمه Hyperparameters ارجاعی از آنها می باشد. شرط ۰.۶ بین اینها می باشد. اینها میتوانند مقداری از این را که در آنها نداشته باشند را افزایش داد.

✓ notebook 2

✓ experiment 1

جواب = ۹۷٪

نمودار

Epoch 2: adam, Adagrad

۷.۸۵

Epoch 3: RMSprop

نمودار

Epoch 21: Adagrad

Epoch 21: adam

۷.۸۲۵۹

Epoch 21: RMSprop

overfitting

Epoch 1: Adagrad

برای این داده ها در آنها از آنها کمترین داده ها و

آخرین داده ای را در آنها نداشته باشند

Fashion MNIST

overfitting

تفویت بحث تاریخ در زبان ای ا و مهد نادر (مسنون) درود ۵۷

1. If you do cross-validation Adversarial experiment  $\rightarrow$  you can't use  
standard regularisation  $\rightarrow$  use a similar Adversarial  
regularization (Notbook p) SGD at zero learning rate  
This will give you patches + (Overfitting) Early stop and  
early weight decay, early stopping in large batches

## Experiment 2

جیسا کوئی نہیں کر سکتا

Exploratory study: Adam et al. (1998) : Robustness of LBN

جـ ٦٨ / دریم نویس یـ گیـس (جـ ۶۷) کـرـدـکـلـهـاـ

~~پیار، مرحوم و ممتاز~~ RMS Prof. Dr. ~~مختار~~ انعامی کوئٹہ

البيانات تشير إلى أن متوسط المدى يزيد قليلاً عن المعيار (std dev = 0.3).

یعنی لا ہوں عملکر آئیں دامت

رسانه آنچه می‌دانیم این است که در مقایسه با آن دیگر روش‌ها، آن روش برای آنکه می‌تواند در هر مرحله از تدریس این روش را بتوانیم بروزگیری کرد، این روش را آنرا در مجموع آن روش می‌دانیم = Adagrad -

ویا تیس ۰.۰۹ یل چینج بزم پروف و ادامه می باشد ۲۳۶ یکمین کشیده  
کنیل یل ۰.۰۹ چینج تیس ۰.۰۹ گریپ لاین ۱۷۰ (۱.۸۸ کیلو)

0.79 Test loss = 0.01 first adam is much more  
efficient (eg 0.41 instead of 0.79) using SGD + momentum

مَنْ هُوَ إِلَّا مَوْلَانَا كَمْ مَنْ هُوَ إِلَّا مَوْلَانَا

أَدَّى إِلَيْهِمْ أَنْتَ وَالْمُؤْمِنُونَ إِذْ أَنْتُمْ تُفْسِدُونَ إِذْ أَنْتُمْ تُفْسِدُونَ

— اس بڑا ہر سوچ درانیں مل نظر پڑے  
ایسا کبھی نہ تین تاریں پین گل کو ملکیں تھیں ۱۸۰۱ء

Chap 14

adam و Capraeas / experiments

(الآن نعم بـ Beta1) Beta1 نعم انتـ

ـ صلـ

ـ لـ Beta1 سـ بـ Beta1 = 0.9 ، Beta1 = 0.95

(epoch ٢٠٠٠ درـ

ـ ٩٩.٥٨٪ عـ Beta1 = 0.95 ، Beta1 = 0.95

ـ كـ

(٩١.٣٪ عـ Beta1 = 0.95)

Beta1 بـ 0.9 نـ وـ Beta1 بـ 0.95 نـ Beta1 بـ 0.95

ـ بـ Beta1 بـ 0.95 اـ Beta1 بـ 0.95

ـ Beta1 بـ 0.95 اـ Beta1 بـ 0.95

ـ سـ

(ـ Beta1 بـ 0.95) Beta1 نـ

ـ صـ

ـ نـ Beta1 بـ 0.95

ـ Beta1 بـ 0.95

ـ بـ

ـ Beta1 بـ 0.95 (0.999 (0.99))

(٩٩.٠٪)

ـ Beta1 بـ 0.95

ـ Beta1 بـ 0.95

تحتی واریانس کو ایکن می بینت، اتفاقاً ایکن واریانس  
 (Variance) کو ایک دیگر نام ایکن فرم (Form) کردا ہے۔  
 مجموعتی ایک زین تک کردا ہے۔  
 (Epsilon (ε) عدالتی (Fairness))

مکانیزم ایجاد این پدیده را می‌توان با در نظر گرفتن تغییرات مخصوصاً در سطح آب و همچنین بررسی میزان پاره شدن این مولکولها در آب می‌توان توضیح داد.

(Robust) ليمونات ديليزاير (DeliZair) هي شركة تابعة لـ Adam للاجهزة الكهربائية  
 $\beta_1 = 0.5$  و  $\beta_2 = 0.9$  و  $\beta_3 = 0.95$  .  
 تمثل الشركة 30% من اسواق الليمونات في مصر .  
 تمثل الشركة 20% من اسواق الليمونات في مصر .  
 $\beta_1 = 0.9$  و  $\beta_2 = 0.95$

## Gemini Chat

- <https://gemini.google.com/share/7e34876ac7cb>  
در این تمرین به طور اشتباه از چت قبلی استفاده کرده ام.

## Notebook LM

- <https://notebooklm.google.com/notebook/70ec6ac4-b5af-4d59-8554-178567f6ea1d>

## References

- <https://arxiv.org/abs/1412.6980>
- <https://www.datacamp.com/tutorial/adam-optimizer-tutorial>
- [https://en.wikipedia.org/wiki/Stochastic\\_gradient\\_descent](https://en.wikipedia.org/wiki/Stochastic_gradient_descent)
- <https://www.kaggle.com/code/tanujasreekanth/comparing-optimizers-fashion-mnist-data-set>
- <https://jamhuri.medium.com/understanding-the-adam-optimization-algorithm-a-deep-dive-into-the-formulas-3ac5fc5b7cd3>
- <https://medium.com/mini-distill/effect-of-batch-size-on-training-dynamics-21c14f7a716e>
- [https://www.reddit.com/r/math/comments/e72q34/the\\_concept\\_of\\_convergence\\_in\\_gradient\\_descent/](https://www.reddit.com/r/math/comments/e72q34/the_concept_of_convergence_in_gradient_descent/)
- <https://sh-tsang.medium.com/brief-review-visualizing-the-loss-landscape-of-neural-nets-dd93cb261afc>