

بسمه تعالی



دانشکده مهندسی کامپیوتر

مهر ۱۴۰۰

یادگیری عمیق

نام استاد: دکتر محمدی

تمرین چهارم

آرمان حیدری

شماره دانشجویی: ۹۷۵۲۱۲۵۲

## ۱. پاسخ سوال اول

الگوریتم‌های بهینه‌سازی adaptive، شامل Adam، RMSprop، AdaGrad، همگی عملکرد بسیار خوبی در مجموعه داده‌های sparse دارند. و همچنین زودتر از الگوریتم‌هایی مانند SGD، momentum SGD، یا Nesterov همگی می‌رسند و learning rate آن‌ها هم کمتر نیاز به بهینه‌سازی و تغییر دارد.

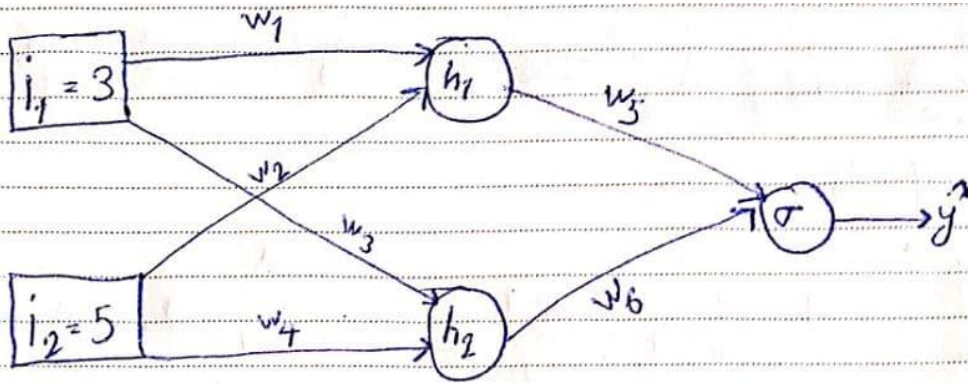
همچنین Adam بهترین آن‌هاست و در مسائل مختلف همگرای زودتر آن‌ها را بر روی داده‌های train ثابت شده است.

علی‌رغم همه این‌ها و استفاده زیاد Adam در مدل‌های امروزه، مقاله‌ها و بحث‌هایی مطرح شده است که الگوریتم‌های SGD، یا Nesterov، یا momentum یا شاید بدون آن‌ها، اگر learning rate مناسبی داشته باشند و به خصوص scheduler خوبی برای نرخ یادگیری داشته باشند می‌توانند بهتر از Adam، generalize کنند. البته با epoch بیشتری به آن نقطه می‌رسند و حتی ممکن است روی داده‌های ~~آموزشی~~ <sup>آزمایشی</sup> هرگز به وقت Adam نرسند اما روی داده‌های جدید می‌توانند بهتر باشند.

در حقیقت اگر مشکلی با زمان آموزش دادن شبکه نداشته باشیم و بتوانیم hyperparam های شبکه SGD را خوب مقدار دهی اولیه و بهینه‌سازی کنیم، می‌توانیم به نتایج بهتری از Adam (که با مقادیر معقولی از ابرپارامترها به نتیجه خوبی می‌رسد) برسیم.

منابع: [medium](#), [shaoanlu](#)

## ۲. پاسخ سوال دوم



epoch=2, loss function="mean square error", weights (vector):

$$\begin{cases} w_1=0.1 \\ w_2=0.2 \\ w_3=0.3 \\ w_4=0.4 \\ w_5=0.5 \\ w_6=0.6 \end{cases}$$

ابتدا forward propagation را اعمال میکنیم. در iteration اول:

$$h_1 = i_1 w_1 + i_2 w_2 = 3 \times 0.1 + 5 \times 0.2 = 1.3$$

$$h_2 = i_1 w_3 + i_2 w_4 = 3 \times 0.3 + 5 \times 0.4 = 2.9$$

$$\hat{y} = \sigma(h_1 w_5 + h_2 w_6) = \sigma(1.3 \times 0.5 + 2.9 \times 0.6) = \sigma(2.39) =$$

$$= \frac{1}{1 + e^{-2.39}} = 0.916$$

تخمین شبکه یادگیری را

محاسبه خطا MSE L:  $\frac{1}{N} \sum_{n=1}^N (y_n - \hat{y}_n)^2$   $\xrightarrow{N=1}$   $(y_n - \hat{y}_n)^2 = (1 - 0.916)^2 = 0.00704$

حالا روابط گرادیان کاهشی را برای هر یک از 6 وزن شبکه باید پیدا کنیم و طبق آن ها آپدیت

$$\frac{dL}{d(y)} = \frac{d(y - \hat{y})^2}{d(\hat{y})} = -2(y - \hat{y}) = 2(\hat{y} - y) = 2(0.916 - 1) = -0.168$$

کنیم:

$$\frac{d\hat{y}}{dw_6} = \frac{d\left(\frac{1}{1 + e^{-h_1 w_5 - h_2 w_6}}\right)}{dw_6} = \frac{d\hat{y}}{d\sigma} \times \frac{d\sigma}{dw_6} = (1 - \sigma)\sigma(h_2) =$$



$$= \left(1 - \frac{1}{1 + e^{h_1 w_5 + h_2 w_6}}\right) \left(\frac{1}{1 + e^{h_1 w_5 + h_2 w_6}}\right) \bar{h}_2 = (1 - 0.916)(0.916) \cdot 2.9$$

$$= 0.2231376$$

$$\textcircled{II} \xrightarrow{\text{با لحاظ گرفتن}} \frac{d\hat{y}}{dw_5} = \left(1 - \frac{1}{1 + e^{h_1 w_5 + h_2 w_6}}\right) \left(\frac{1}{1 + e^{h_1 w_5 + h_2 w_6}}\right) h_1 = 0.1000272 \quad \textcircled{III}$$

$$\textcircled{II} \xrightarrow{\text{با لحاظ گرفتن}} \frac{d\hat{y}}{dh_2} = \left(1 - \frac{1}{1 + e^{h_1 w_5 + h_2 w_6}}\right) \left(\frac{1}{1 + e^{h_1 w_5 + h_2 w_6}}\right) w_6 = 0.0384464 \quad \textcircled{IV}$$

$$\textcircled{III} \xrightarrow{\text{با لحاظ گرفتن}} \frac{d\hat{y}}{dh_1} = \left(1 - \frac{1}{1 + e^{h_1 w_5 + h_2 w_6}}\right) \left(\frac{1}{1 + e^{h_1 w_5 + h_2 w_6}}\right) w_5 = 0.0461356 \quad \textcircled{V}$$

$$\frac{dh_1}{dw_1} = i_1 = 3, \quad \frac{dh_1}{dw_2} = i_2 = 5$$

$$\frac{dh_2}{dw_3} = i_1 = 3, \quad \frac{dh_2}{dw_4} = i_2 = 5$$

حال با استفاده از قانون مشتق زنجیره‌ای، روابط مستقیم گراپین زیر را می‌توانیم:

$$\frac{dL}{dw_1} = \frac{dL}{d\hat{y}} \times \frac{d\hat{y}}{dh_1} \times \frac{dh_1}{dw_1} = 2(\hat{y} - y) \times \sigma \times (1 - \sigma) \times w_5 \times i_1$$

$$\xrightarrow{\text{با لحاظ گرفتن}} \frac{dL}{dw_2} = 2(\hat{y} - y) \times \sigma \times (1 - \sigma) \times w_5 \times i_2$$

$$\xrightarrow{\text{با لحاظ گرفتن}} \frac{dL}{dw_3} = 2(\hat{y} - y) \times \sigma \times (1 - \sigma) \times w_6 \times i_1$$

$$\xrightarrow{\text{با لحاظ گرفتن}} \frac{dL}{dw_4} = 2(\hat{y} - y) \times \sigma \times (1 - \sigma) \times w_6 \times i_2$$

$$\frac{dL}{dw_5} = \frac{dL}{d\hat{y}} \times \frac{d\hat{y}}{dw_5} = 2(\hat{y} - y) \times \sigma(1 - \sigma) \times h_4$$

$$\frac{dL}{dw_6} = \frac{dL}{d\hat{y}} \times \frac{d\hat{y}}{dw_6} = 2(\hat{y} - y) \times \sigma(1 - \sigma) \times h_2$$

الغرض روابط گزاردیان را دقیق به دست آورده ایم. و وقت آپدیت کردن وزن ها است:

مثال 6.1.1:

$$w_i = w_i - (\text{learning rate}) \times \frac{dL}{dw_i}$$

با نرخ learning = 0.1  
rate

$$w_1 = 0.1 - 0.1 \times (-0.0193627) = 0.10193627$$

$$w_2 = 0.2 - 0.1 \times (-0.032271) = 0.20322712$$

$$w_3 = 0.3 - 0.1 \times (-0.023235) = 0.30232353$$

$$w_4 = 0.4 - 0.1 \times (-0.0387255) = 0.40387255$$

$$w_5 = 0.5 - 0.1 \times (-0.0167811) = 0.50167811$$

$$w_6 = 0.6 - 0.1 \times (-0.0374347) = 0.61874347$$

حال به پایان اولین epoch رسیدیم و این مراحل را دوباره با وزن های جدید تکرار می کنیم.

روابط دقیقاً مثل قبل هستند و لذا فقط نتایج که با ماشین حساب محاسبه کردیم را می نویسیم:

$$\hat{y} = 0.9229156 \quad \text{تقریب شده بین از دو بار تکرار}$$



وزن‌های شبکه پس از  
محاسبه گرادینان  
و آپدیت آن‌ها

$$w_1 = 0.1036090$$

$$w_2 = 0.2060151$$

$$w_3 = 0.3043593$$

$$w_4 = 0.4072655$$

$$w_5 = 0.5098404$$

$$w_6 = 0.6347652$$

شبکه هگزا نشده است. چون بار دوم اجرا مقادیر گرادینان تقریباً مشابه بار اول بدست آمد و وزن‌ها همچنان در همان جهت قبلی آپدیت می‌شوند. اگر  $learning\ rate$  را زیاد کنیم، یا از تابع ضرری استفاده کنیم که خطای بیشتری بازدارنده ممکن است بتوانیم در 2 تکرار به جواب برسیم.

### ۳. پاسخ سوال سوم

کد و توضیح مرحله به مرحله آن در فایل HW4.ipynb، ([لینک گوگل کولب](#)) section = Question 3 آمده است.

در حالتی که `data_augmentation = True` باشد، شبکه تعداد رندومی عکس جدید که شامل چرخش ۴۵ درجه یا افقی، یا کراپ شدن عکس های قبلی باشد میسازد و آن ها را هم به دیتا اضافه میکند (با همان label عکس اصلی). که این کار باعث robust شدن شبکه می شود.

خروجی دقت های شبکه ما روی داده آموزشی برای هر کدام از کلاس های این دیتاست به شکل زیر بود:

```
airplane train accuracy: 77.42
automobile train accuracy: 79.88
bird train accuracy: 41.32
cat train accuracy: 42.0
deer train accuracy: 41.78
dog train accuracy: 48.1
frog train accuracy: 59.86
horse train accuracy: 68.34
ship train accuracy: 50.84
truck train accuracy: 67.16
```

و برای داده های تست هم به شکل زیر به دست آمد:

```
airplane test accuracy: 77.7
automobile test accuracy: 75.6
bird test accuracy: 40.6
cat test accuracy: 38.4
deer test accuracy: 40.4
dog test accuracy: 48.0
frog test accuracy: 60.4
horse test accuracy: 65.9
ship test accuracy: 45.9
truck test accuracy: 65.5
```

البته دو بار تکرار برای همگرایی شبکه های عمیق عدد بسیار کمی است. اما در همین مدت هم میبینیم که دقت ها از عدد احتمالاتی که باید به دست بیاید (حدودا ۱۰ درصد برای هر کلاس) بسیار بیشتر است و مشخص است که

شبکه آموخته است. دقت داده تست هم مطابق انتظار در تمام کلاس ها از آموزشی کمتر است، اما در حدود داده آموزشی همان کلاس خاص است.