

زمان باقیمانده 2:44:20

الف) اگر در یک شبکه کانولوشنی عمل کانولوشن با افزودن صفر در مرزها انجام شود و بعد ورودی  $106 \times 64$ ، بعد فیلتر  $5 \times 5$  و اندازه گام حرکت 3 باشد، اندازه تصویر خروجی چند در چند خواهد بود؟ (ب) افزودن صفر به چه منظور انجام میشود؟ (7.5 نمره)

سؤال 1

هنوز پاسخ داده نشده است

نمره از 1.00

سؤال 1

input size =  $106 \times 64$   
Filter size =  $5 \times 5$   
Strid = 3

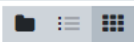
output size =  $((N-F)/S) + 1 \Rightarrow$   
 $\Rightarrow ((106-5)/3) + 1 = 34$  برای عدد  
 $\Rightarrow ((64-5)/3) + 1 = 19$  برای عدد

افزودن 0 (Zero padding) به منظور جلوگیری از کم شدن اندازه داده ها بر اثر padding است و نتیجه برآورد و تمام یکسایزها قرار بگیرد عیناً مرکز فیلتر روی بیشتر داده ها قرار بگیرد، بنابراین پیرامون داده 0 اضافه میکنیم.

زمان باقیمانده 2:43:57

کانولوشن  $1 \times 1$  روی حجمی از داده های ورودی، چه اثری روی داده ها ایجاد میکند؟ (2.5 نمره)

حداکثر اندازه فایل های جدید: 1 گیگابایت



فایل ها

سؤال 2

هنوز پاسخ داده نشده است

نمره از 1.00

علامت زدن

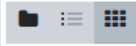
سؤال 2

کانولوشن  $1 \times 1$  حجم داده ها را تعداد کانال ها را کاهش میدهد. مثلاً اگر داده ها  $28 \times 28 \times 72$  باشد با 21 فیلتر  $1 \times 1 \times 72$  ابعاد داده به صورت  $28 \times 28 \times 21$  در می آید. با این کار اطلاعات ورودی از بین نمی رود بلکه داده های آن فشرده تر میشوند و این کار باعث کمتر شدن حجم محاسبات می شود.

زمان باقیمانده 2:43:49

چرا در بسیاری از کاربردها نرمالسازی داده ها ضروری است؟ در مسائل دسته بندی نرمال سازی روی چه داده هائی انجام میشود؟ (5 نمره)

حداکثر اندازه فایل های جدید: 1 گیگابایت



سؤال 3

هنوز پاسخ داده نشده است

نمره از 1.00

۳ علامت زدن سؤال

سؤال 3

نرمال سازی داده ها به دلایل زیر انجام می شود:

۱- از بین بردن Correlation بین Variable ها

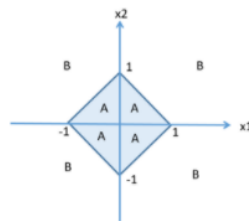
۲- از بین بردن داده های redundant

۳- برخی Variable ها رنج بزرگتری دارند و به همین خاطر بایاس در نتایج را ایجاد می کنند. به این منظور همه داده ها باید در یک scale قرار داشته باشند.

۴- متوسط نرمال سازی می توانیم داده های غیر استاندارد را حذف کنیم > Trimming Winsorizing

زمان باقیمانده 2:43:34

میخواهیم مسئله دسته بندی دو دسته ای زیر را با یک شبکه جلو رو حل کنیم. (ا) معماری شبکه مناسب برای این کار را مشخص کنید. (ب) وزنه های شبکه را بدست آورید. (۲۰ نمره)



سؤال 4

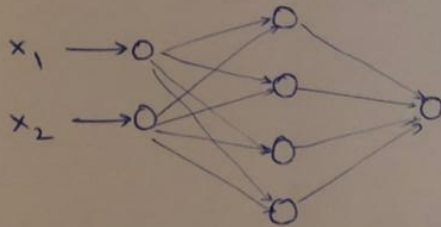
هنوز پاسخ داده نشده است

نمره از 1.00

۳ علامت زدن سؤال

سؤال 4

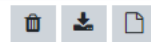
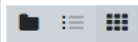
برای انجام این دسته بندی می توانیم از یک MLP استفاده کنیم. که تعداد اصلاعی برابر تعداد ورودی ها در لایه مخفی خواهد بود.



زمان باقیمانده 2:43:13

در شبکه مولد تقابلی، چه چیز باعث میشود که بخش مولد بتواند یاد بگیرد نمونه های مشابه نمونه های واقعی تولید کند؟ (2.5 نمره)

حداکثر اندازه فایل های جدید: 1 گیگابایت



سؤال 5

هنوز پاسخ داده نشده است

نمره از 1.00

۳ علامت زدن سؤال

سؤال 15

~~بخش Generator~~  
مولد سعی دارد که داده های تولید شده را با آنهایی که مقایسه کننده را فریب دهد و به همین خاطر این داده ها باید شبیه به داده های واقعی باشند به طوری که نتوانیم فریب مقایسه کننده برای داده اصلی هدف در GAN:

$$\min_{\theta_g} \max_{\theta_d} \left[ E_{x \sim p_{data}} \log D_{\theta_d}(x) + E_{z \sim p(z)} \log (1 - D_{\theta_d}(G_{\theta_g}(z))) \right]$$

generator parameters      discriminator parameters      فریب مقایسه کننده برای داده تولید شده

ابتدا Generator غیر ~~تولید شده~~ را دریافت می کنیم و سعی می کنیم  $D_{\theta_d}(G_{\theta_g}(z))$  را  $\max$  کنیم چون فریب مقایسه کننده برای داده اصلی 1 است و برای این کار هم باید احتمال این که مقایسه کننده در اشتباه است را بالا ببریم  
 $\max_{\theta_g} E_{z \sim p(z)} \log (D_{\theta_d}(G_{\theta_g}(z)))$   
 برای مثال Generator می تواند لایه های deconvolution داشته باشد.

زمان باقیمانده 2:42:58

در خودکدگذار تغییراتی، برای آنکه هر مؤلفه بردار پنهان مستقلاً یک خصیصه داده را ارائه دهد، چه شرطی باید برقرار باشد؟  
(5 نمره)

حداکثر اندازه فایل‌های جدید: 1 گیگابایت

سؤال 6

هنوز پاسخ داده نشده است

نمره از 1.00

۳ علامت زدن سؤال

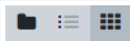
سؤال 6

برای آن که هر مؤلفه بردار پنهان  $(h)$  یک ویژگی مجرد را تولید کند باید  
توزیع احتمالات مؤلفه‌های بردار پنهان به نحوی باشد که همبستگی بین آن‌ها  
صفر نداشته باشد.

زمان باقیمانده 2:42:38

الف) حفظ توپولوژی در شبکه‌های ویژگی خود سازمانده به چه معنی است؟ ب) در این شبکه چگونه این خصیصه ایجاد شده است؟ (5 نمره)

حداکثر اندازه فایل‌های جدید: 1 گیگابایت



فایل‌ها

سؤال 7

هنوز پاسخ داده نشده است

نمره از 1.00

۳ علامت زدن سؤال

سؤال 7

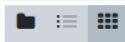
مقطع تصویربرداری به این معناست که وقتی داده‌های فضای کم بعدتر نگاشت داده  
می‌شوند ویژگی‌های داده‌ها فقط شود (یعنی اگر در فضای اول داده‌ها هم نزدیک  
همیشه در فضای جدید هم به هم نزدیک باشند). بنابراین نگاشت به گونه‌ای انجام  
می‌شود که ارتباط بین داده‌ها فقط شود.

در شبکه‌های عمیق از طریق ایجاد رقابت (برای شبکه‌های رقابتی) این کار انجام می‌شود.  
در این شبکه برای هر نورون یک صایه تعریف می‌شود و در هنگام اصلاح وزن  
هر نورون صایه‌های آن هم اصلاح وزن دارند (بهرجهت به میزان نزدیکی به نورون  
اصلی). با این کار یک نامیه به دورن حاس می‌شود و فقط یک نورون

زمان باقیمانده 2:42:24

الف) آیا ضمانتی وجود دارد که در خودکدگذارها خروجی کدگذار حتما ویژگیهای داده‌های ورودی را یادگرفته باشد؟ (ب) اگر  
جواب منفی است آیا میتوان از طریقی این ضمانت را ایجاد کرد؟ توضیح دهید. (7.5 نمره)

حداکثر اندازه فایل‌های جدید: 1 گیگابایت



فایل‌ها

سؤال 8

هنوز پاسخ داده  
نشده است

نمره از 1.00

علامت زدن  
سؤال

سؤال 8

در هر گذارها  $h$  overcomplete که به معنای این است که ابعاد ورودی  $X$  بزرگتر است از ابعاد  $h$  است. ~~این به معنای این است که ابعاد ورودی  $X$  بزرگتر است از ابعاد  $h$  است.~~ که در این حالت  $X$  را کمترین مقدار  $h$  به صورتی که بتواند  $X$  را بازسازی کند انتخاب می‌کنیم.

(1) regularize autoencoder، مثل weight decay به تابع هدف اضافه می‌شود و باعث می‌شود وزن‌ها کوچک‌تر شوند.

(2) sparse autoencoder: اتوالی که خروجی تولید می‌کند اسپارس است.

$$J_{SAE}(\theta) = \sum_{x \in D_n} L(x, g(f(x))) + \eta \sum_i KL(p || p_i)$$

$p_i$ : میانگین خروجی برای ورودی  $x$  در  $m$  داده ورودی است.

$p$ : یک میانگین مرکب است (0.05)

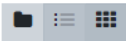
به سبب می‌شود خروجی ورودی  $p$  به سمت 0 برود و در نتیجه خروجی فرض ورودی  $h$  می‌شود.

(3) denoising autoencoder: داده‌ها را نویز می‌کنیم و به اتوالی می‌دهیم و AE باید سعی کند نویزها را بدون نویز برگرداند.

زمان باقیمانده 2:42:11

چرا در بعضی شبکه‌ها برای هر لایه ورودی لایه را با خروجی آن ترکیب می‌کنند و به لایه بعدی می‌دهند؟ این کار چه تأثیری دارد؟ (2.5 نمره)

حداکثر اندازه فایل‌های جدید: 1 گیگابایت



فایل‌ها

سؤال 9

هنوز پاسخ داده نشده است

نمره از 1.00

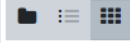
3 علامت زدن سؤال



زمان باقیمانده 2:41:57

در آموزش شبکه های جلورو با استفاده از تنظیم، توضیح دهید که تنظیم بصورت L2 و L1 هرکدام چگونه در آموزش تاثیر میگذارند. (7.5 نمره)

حداکثر اندازه فایل های جدید: 1 گیگابایت



فایل ها

سؤال 10

هنوز پاسخ داده نشده است

نمره از 1.00

علامت زدن سؤال

سوال 10

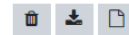
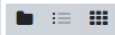
در منظم ساز  $L_1$  مجموع مقادیر وزن ها را به تابع هزینه اضافه می کنیم  
( $\sum_{i=1}^n |w_i|$ ) که از  $overfitting$  جلوگیری کند از جمله فنی های این  
منظم ساز این است که نسبت به داده های پرت قوی است اما نمی تواند  
انگیزه های پیچیده را یاد بگیرد (همچنین اسپارس تر می باشد) و چون اسپارس است  
در انتخاب  $feature$  می تواند مدور استفاده قرار گیرد.

در  $L_2$  ( $\sum_{i=1}^n |w_i|^2$ ) مجموع مربع وزن ها به تابع اضافه می شود.  
این روش اسپارس نیست (یعنی هیچ وزن 0 نمی شود) و ضرایب کوچک می شوند  
(که این برای همه ضرایب است) و از این رو در برابر داده های پرت قوی  
نسبت به وزن از آن استفاده می کنیم کمترین ویژگی ها  $Correlation$  را  
ببینیم

زمان باقیمانده 2:41:42

تفاوت های بین سلول های GRU و LSTM را لیست کنید. (7.5 نمره)

حداکثر اندازه فایل های جدید: 1 گیگابایت



سؤال 11

هنوز پاسخ داده نشده است

نمره از 1.00

علامت زدن سؤال

## سؤال 11

تفاوت اصلی بین LSTM و GRU

1) GRU دو گیت دارد (reset و update) در حالی که LSTM

3 گیت دارد (input, output, forget)

2) GRU از نظر محاسبات efficient تر است.

3) GRU مبرین اطلاعات را کنترل می کند بدون این که واحد memory راسته باشد.

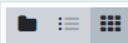
4) آموزش GRU سریع تر است.

5) با توجه به این که ساختار GRU ساده تر است می تواند آن را در دستگاه های کم توان اجرا کند.

زمان باقیمانده 2:41:19

میخواهیم برای پیش بینی برق مصرفی شبکه مترو در یک شهر مدلی پیشنهاد دهیم. پیشنهاد شما چیست؟ چه جنبه هایی باید در طراحی این مدل در نظر باشد؟ (10 نمره)

حداکثر اندازه فایل های جدید: 1 گیگابایت



[فایل ها](#)

سؤال 12

هنوز پاسخ داده نشده است

نمره از 1.00

3 علامت زدن سؤال



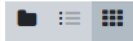
## سؤال 12

مصرف برق در ساعت‌های مختلف متفاوت است، هم‌مینه مصرف برق در روزها و مختلف هفته و هم‌مینه فصل‌های مختلف سال نیز متفاوت می‌باشد و طراحی یک مدل برای پیش‌بینی داشتن این مقیاس‌ها دشوار است بنابراین داده‌ها را فواید می‌کنیم ← هر ساعت هر روز در یک فواید قرار می‌گیریم بنابراین هر فواید load demand برای یک ساعت مشخص است. پس 24 فواید خواهیم داشت. و برای هر فواید Fully recurrent Network استفاده می‌شود. و برای پیش‌بینی یک روز در یک ساعت مشخص، دیتای مربوط به روز جاری و چند روز قبل از آن ساعت بعنوان در ورودی گرفته می‌شوند.

زمان باقیمانده 2:41:11

یک توصیه بسیار مثبت در آموزش شبکه‌های جلودر برای پیش‌بینی سری‌های زمانی چیست؟ (2.5 نمره)

حداکثر اندازه فایل‌های جدید: 1 گیگابایت



فایل‌ها

## سؤال 13

هنوز پاسخ داده نشده است  
نمره از 1.00  
علامت زدن سؤال

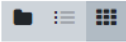
## سؤال 13

- ۱) اکثر برای آموزش شبکه از کل داده‌ها استفاده کنیم، آموزش مبتنی بر خواص را داریم
- ۲) هم‌مینه مبتنی بر خواص علاوه بر داده‌های سری‌زمانی مستقیم که می‌خواهیم آن را پیش‌بینی کنیم، داده‌هایی مربوط به مستقیم‌ها را نیز به شبکه می‌دهیم. زیرا پیش‌بینی یک مستقیم به مستقیم‌ها را نیز وابسته است.
- ۳) ؟ داده‌های جدیدتر (more recent) وزن بیشتری می‌گیریم.

زمان باقیمانده 2:40:53

در روش گرادینان خط مشی در یادگیری تقویتی عمیق، توضیح دهید که شبکه چگونه آموزش داده میشود. (5 نمره)

حداکثر اندازه فایل‌های جدید: 1 گیگابایت



فایل‌ها

سؤال 14

هنوز پاسخ داده نشده است

نمره از 1.00

علامت زدن سؤال

سؤال 14

Deep policy based RL :

در این روش منافعیم شبکه حالت را بگیرد و فریب های مکرر اینها را کندکند این فریب ها هر کدام احتمال این که عمل عینیه است را می دهد.

$\sum_{a_i \in A} \pi(a_i|s) = 1$

$\pi(a|s) = p(\text{action}|\text{state})$

آموزش :

- 1) یک سیاست را برای مدیریت اعمال کند
- 2) وقت یاداشن بالا برود احتمال را بالا ببرند و برعکس.
- 3) بعد از مکرر تکرار نشان منجر به برنده شدن می شود.

Function reinforce

```
initialize  $\theta$ 
for episode  $\sim \pi_\theta$ 
   $\{s_i, a_i, r_i\}_{i=1}^{T-1} \leftarrow \text{episode}$ 
  for  $t = 1$  to  $T-1$ 
     $\nabla \leftarrow \nabla_{\theta} \log \pi_{\theta}(a_t | s_t) R_t$ 
     $\theta \leftarrow \theta + \alpha \nabla$ 
  return  $\theta$ 
```

$\log \text{likelihood} + R_t \text{ actions}$

$\nabla_{\theta} \log \pi_{\theta}(a_t | s_t) R_t$

$\log \text{likelihood} \quad \text{reward}$

$= \text{action}$

زمان باقیمانده 2:40:45

سؤال 15

### هنوز پاسخ داده

نشده است

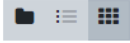
نمره از 1.00

۱۲ علامت زدن

### سؤال

در شبکه کانولوشنی، لایه های ادغام چه کارهائی انجام میدهند؟ شرح دهید که هر مورد با ادغام چگونه انجام میشود. (10 نمره)

حداکثر اندازه فایل‌های جدید: 1 گیگابایت



فایل‌ها

بیمیه ها کانونی هستند شامل همه یا بعضی از عدول زیر می باشند:

- Convolution
- Nonlinearity
- Spatial pooling
- Normalization

}  $\Rightarrow$  Convolution layer

}  $\Rightarrow$  pooling layer

در لایه های ارتفاع مکانی: این صورت است که در سفت‌ترین نقشه همان ابعاد شبکه توسط کانولوشن من خواهم داده ها را در حجم ارتفاع کنیم و حجم داده ها را کاهش دهیم. مثلاً اگر 50 فیلتر اعمال کنیم 50 صفحه دیگر خواهیم داشت که در لایه در درون لایه دیگر است. این به شدت با حجم زیاد از داده ها مواجه هستیم که باید محدود کنترل شود. بنابراین در نقشه زیر نامیه های را در نظر میگیریم و  $sum \leq average \leq T_{max}$  زیر نامیه را در نظر میگیریم. بر اساس stride که بگیریم می‌کنیم که می‌تواند تا  $F$  (ساند فیلتر) باشد. نامیه ها می‌توانند همپوشانی داشته و یا نداشته باشند. spatial pooling علاوه بر این که حجم داده ها کم می‌شود باعث می‌شود که نامیه ها در نهایت در لایه بعد بزرگتر شود و در درون همان شبکه را ببینیم. در عقب نرمال سازی هم (contrast normalization) که می‌تواند قبل از ارتفاع یا بعد از آن باشد، دامنه ها را نرمال کرده که در یادگیری در شبکه دارند.