

# به نام آنکه جان را فکرت آموخت



## دانشگاه صنعتی شریف دانشکده مهندسی برق

### آزمون میان نیم سال، سال تحصیلی ۱۴۰۰-۰۱ درس یادگیری عمیق

نام و نام خانوادگی دانشجو:

شماره دانشجویی: .....

#### نکات مهم:

- آزمون کتاب و جزوه بسته است.
- آزمون شامل ۸ سوال و سه مسئله و مدت آن ۱۲۰+۱۵ دقیقه است.
- استفاده از دو برگ کاغذ A4 مجاز است.
- نام فایلی که آپلود می کنید فقط شما دانشجویی شما باشد (بدون هیچ اضافاتی): **98123456.pdf**
- مسیر ارسال پاسخنامه فقط **CW** است، از ارسال با ایمیل یا واتساپ خودداری فرمایید.
- لطفا پاسخنامه را با قلم سیاه یا آبی پررنگ بر روی کاغذ سفید بنویسید.

به سوالات داده شده در فضای زیر آنها پاسخ دهید: (هر یک ۱۰ نمره)

۱) یک شبکه عصبی عمیق در تنها سه بار مقدار دهی اولیه تصادفی با خطای بالایی آموزش می‌بیند، در مورد پیچیدگی شبکه چه می‌توان گفت؟ با توجه به تعداد اندک مقداردهی اولیه، ممکن است شبکه پیچیده بوده (با توجه کلمه عمیق در متن سوال این حالت وجود دارد) و در مینیمم محلی یا نقاط زینی گیر افتاده باشد، البته حالت بسیار ساده ساختار نیز محتمل است که موضوع این سوال نیست (کلمه عمیق در متن).

۲) آیا تابع فعالیت softmax نسبت به افزودن یک مقدار ثابت به ورودی آن حساس می‌باشد. (منظور مقایسه بین  $\text{softmax}(z)$  و  $\text{softmax}(z+C)$  می‌باشد). خیر حساس نیست با نوشتن رابطه، ساده‌سازی صورت و مخرج رخ می‌دهد (نوشتن روابط ضروری است)

۳) تعدادی داده در فضای دو بعدی که بطور تقریبی از نمودار  $x_2 = (x_1)^2$  تبعیت می‌کنند را به دو روش زیر مدل می‌کنیم:

الف) یک خط  $x_2 = ax_1 + b$ : بایاس بزرگ و واریانس کوچک

ب) رسم یک پاره خط از هر نقطه به نقطه بعدی: بایاس کوچک و واریانس بزرگ  
برای هر مدل در مورد بایاس و واریانس مدل چه می‌توان گفت (با استفاده از کلمات بزرگ و کوچک در مورد بایاس و واریانس)

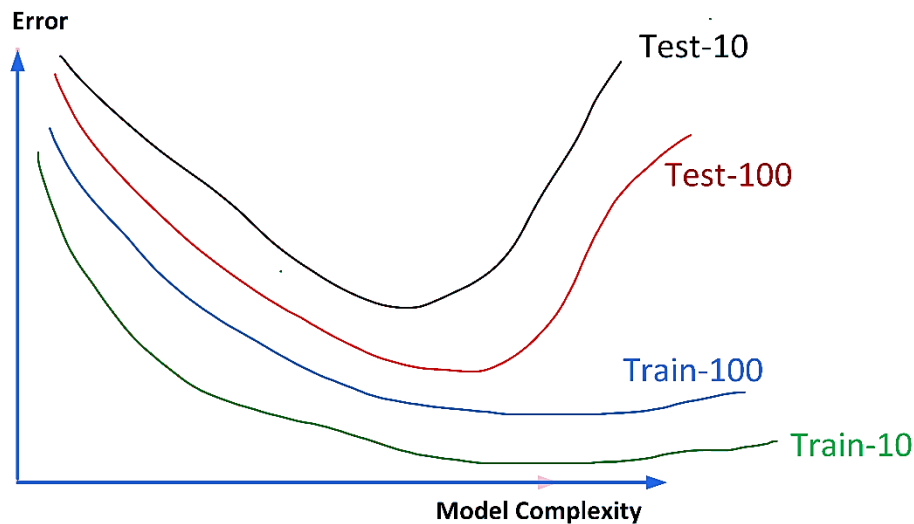
۴) یک شبکه عمیق را با تقسیم داده‌های آموزش و آزمایش به نسبت ۸۵:۱۵ درصد برای آزمایش و ۸۵ درصد برای آموزش) و در دو تجربه متفاوت در دو حالت ۱۰,۰۰۰ داده (در مجموع) و ۱۰۰,۰۰۰ داده (در مجموع) آموزش می‌دهیم. (هر دون دسته داده از یک توزیع آماری تبعیت می‌کنند) با ذکر دلیل نمودار خطای هر یک از ۴ حالت زیر را بر حسب پیچیدگی مدل رسم نمایید:

الف) خطای آموزش در حالت استفاده از ۱۰,۰۰۰ داده. برچسب نمودار: Train-10  
در اینحالت با توجه به تعداد داده کم (تنوع کمتر نسبت به حالت ۱۰۰,۰۰۰) خطای آموزش کم است.

ب) خطای آزمایش در حالت استفاده از ۱۰,۰۰۰ داده. برچسب نمودار: Test-10  
در اینحالت تعمیم پذیری شبکه کم است (تعداد داده کمتر در آموزش) و خطای آزمایش بالا است.

الف) خطای آموزش در حالت استفاده از ۱۰۰,۰۰۰ داده. برچسب نمودار: Train-100  
در اینحالت با توجه به تعداد داده بالا و تنوع بالا در داده موجود، خطای آموزش بطور نسبی بالاتر خواهد بود.

الف) خطای آزمایش در حالت استفاده از ۱۰۰,۰۰۰ داده. برچسب نمودار: Test-100  
در اینحالت تعمیم پذیری شبکه بالا است (تعداد داده بیشتر در آموزش) و خطای آزمایش بطور نسبی کمتر است.



(۵) لایه دوم یک شبکه CNN دارای ابعاد  $12 \times 200 \times 200$  است، چرا در فرآیند آموزش ضرایب ۱۲ فیلتر این لایه مشابه هم نمی‌شوند؟ به دلیل مقدار دهی تصادفی اولیه متفاوت

(۶) تفاوت Unpooling و DeConvLayer را به اختصار شرح دهید. قابلیت یادگیری ضرایب که در DeConvLayer وجود دارد

(۷) لایه Spatial Pyramid Pooling چه ویژگی خاصی دارد؟ (به اختصار شرح دهید): فیکس کردن سائز ورودی قبل از لایه Dense و استفاده از رزولوشن‌های مختلف با توجه به تقسیمات نه بر اساس تعداد پیکسل بلکه بر اساس سطح تقسیمات

(۸) برای تقریب تکه‌ای خطی توابع یک بعدی- یک متغیری چه معماری از شبکه عصبی پیشنهاد می‌کنید. (ذکر تعداد لایه‌ها و نوع توابع فعالیت ضروری است)

Number of input: **One (1)**

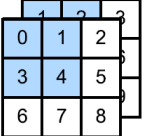
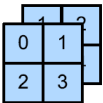


Number of output: **One (1)**

Number of Hidden Layer:  **$\geq$  One ( $\geq 1$ )**

Hidden Layer Activation Function: **ReLu**

Output Layer Activation Function: **Linear**

مسئله اول (۲۰ نمره): در لایه کانولوشنی زیر ماتریس خروجی را محاسبه و تکمیل نمایید. (با فرض حالت Valid)

Input	Kernel	Input	Kernel	Output
				$\begin{bmatrix} 56 & 72 \\ 104 & 120 \end{bmatrix}$

مسئله دوم (۳۰ نمره): معادلات حاکم بر یک شبکه عصبی فرضی با دو ورودی تصویری (که برداری شده‌اند) به شرح زیر است:

$$\left. \begin{array}{l} x_1, x_2 \in \mathbb{R}^n \\ W \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m \end{array} \right| \begin{array}{l} h_1 = \text{sigmoid}(Wx_1 + b) \\ h_2 = \text{sigmoid}(Wx_2 + b) \end{array} \left| \begin{array}{l} J = 0.5 \|h_1 - h_2\|^2 + 0.5 \lambda \|W\|_F^2 \\ \|A\|_F^2 = \sum_{i,j} |A_{i,j}|^2 \end{array} \right.$$

الف) این شبکه چه عملی انجام می‌دهد؟ وظیفه اصلی این شبکه یادگیری یک معیار شباهت غیرخطی بین دو ورودی است که می‌تواند در تصویر یا حتی wordembedding کاربرد داشته باشد. که در هنگام آموزش، برای ورودی‌های مشابه، خروجی اندک تولید کند و برای ورودی‌های متفاوت خروجی بزرگ تولید کند.

پ) معادلات آموزش بردارهای وزن و بایاس ورودی را بنویسید (با استفاده یک SGD ساده با سایز Batch برابر ۱۰)

$$z_1 = Wx_1 + b, \quad z_2 = Wx_2 + b$$

$$\Delta_1 = \frac{\partial J}{\partial h_1} = h_1 - h_2, \quad \Delta_2 = \frac{\partial J}{\partial h_2} = h_2 - h_1$$

$$\Delta_3 = \frac{\partial J}{\partial z_1} = \Delta_1 \odot h_1 \odot (1 - h_1), \quad \Delta_4 = \frac{\partial J}{\partial z_2} = \Delta_2 \odot h_2 \odot (1 - h_2)$$

$$\frac{\partial J}{\partial W} = \Delta_3 x_1^T + \Delta_4 x_2^T + \lambda W, \quad \frac{\partial J}{\partial b} = \Delta_3 + \Delta_4$$

$$W_{t+1} = W_t - \varepsilon \frac{1}{10} \sum_{m=1}^{10} \frac{\partial J_m}{\partial W}, \quad b_{t+1} = b_t - \varepsilon \frac{1}{10} \sum_{m=1}^{10} \frac{\partial J_m}{\partial b}$$

مسئله سوم (۴۰ نمره): تابع هدف ساده شده یک مدل یادگیری به شکل  $0.5\omega^T H \omega$ ، اگر تجربه مقدار-بردار ویژه زیر را انجام دهیم:

$$H = Q\Lambda Q^T$$

الف) فرمول یادگیری ضریب  $\omega$  را با روش معمول (Gradient Descent) با فرض طول گام  $\varepsilon$  بنویسید.

$$w_{t+1} = w_t - \alpha \nabla_w (0.5 w_t^T H w_t) = w_t - \alpha H w_t = (I - \alpha H) w_t = (I - \alpha Q \Lambda Q^T) w_t$$

$$\xrightarrow{QQ^T=I} w_{t+1} = Q(I - \alpha \Lambda) Q^T w_t$$

ب) با شروع از شرط اولیه  $\omega_0$  مقدار این ضریب را گام شماره  $t$  بدست آورید.

$$w_1 = Q(I - \alpha \Lambda) Q^T w_0 \Rightarrow w_2 = Q(I - \alpha \Lambda) Q^T w_1 = Q(I - \alpha \Lambda) Q^T Q(I - \alpha \Lambda) Q^T w_0 = \dots$$

$$= Q(I - \alpha \Lambda)^2 Q^T w_0$$

$$\therefore w_t = Q(I - \alpha \Lambda)^t Q^T w_0$$

پ) تحت چه شرطی این الگوریتم همگرا می شود؟

گلوگاه همگرایی درایه های قطر اصلی ماتریس قطری  $(I - \alpha \Lambda) = \text{diag}(1 - \alpha \lambda_i)$  لذا باید:

$$|1 - \alpha \lambda_{MAX}| < 1 \Rightarrow \underset{>0}{-1} < \underset{>0}{1 - \alpha \lambda_{MAX}} < 1 \Rightarrow \underset{>0}{-2} < \underset{>0}{-\alpha \lambda_{MAX}} < 0 \Rightarrow \alpha < \frac{2}{\lambda_{MAX}}$$