

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؟

(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).

(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 $x_2 = ax_1 + b$

ب) رسم یک پاره خط از هر نقطه به نقطه بعدی

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 (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**

ب) خطای آزمایش در حالت استفاده از ۱۰,۰۰۰ داده. برچسب نمودار: **Test-10**

الف) خطای آموزش در حالت استفاده از ۱۰۰,۰۰۰ داده. برچسب نمودار: **Train-100**

الف) خطای آزمایش در حالت استفاده از ۱۰۰,۰۰۰ داده. برچسب نمودار: **Test-100**

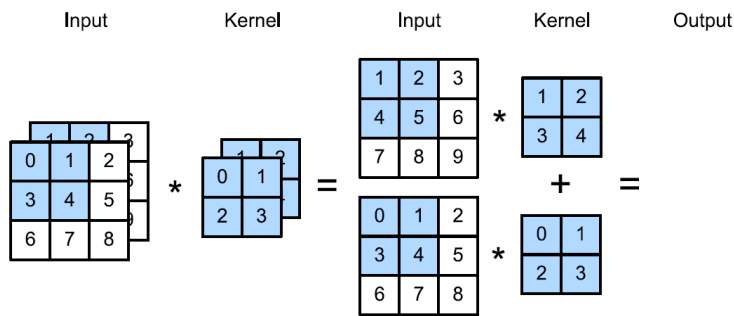
(۵) لایه دوم یک شبکه CNN دارای ابعاد $200 \times 200 \times 12$ است، چرا در فرآیند آموزش ضرایب ۱۲ فیلتر این لایه مشابه هم نمی‌شوند؟

(۶) تفاوت Unpooling و DeConvLayer را به اختصار شرح دهید.

(۷) لایه Spatial Pyramid Pooling چه ویژگی خاصی دارد؟ (به اختصار شرح دهید)

(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)

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



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؟

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

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 ω را با روش معمول (Gradient Descent) با فرض طول گام ε بنویسید.

ب) با شروع از شرط اولیه ω_0 مقدار این ضریب را گام شماره t بدست آورید.

پ) تحت چه شرطی این الگوریتم همگرا می‌شود؟