

نام و نام خانوادگی:

شماره دانشجویی:

۱. یک داده ورودی از اندازه (۱۰, ۱۰, ۳) را در نظر بگیرید که به یکی از دو لایه زیر داده می شود:
أ. لایه تماماً متصل (fully-connected) شامل دو نرون با ورودی بایاس.
ب. لایه پیچشی (convolutional) با سه فیلتر ۲x۲, zero-padding و گام حرکت ۲ (stride=۲).
فرض کنید که ورودی به لایه تماماً متصل ابتدا به صورت یک بردار یک بعدی تغییر شکل یابد. تعداد پارامترهای قابل آموزش در هر یک از این دو لایه چقدر است؟ اگر لایه کانولوشنی انتخاب گردد، ابعاد خروجی آن چقدر خواهد بود؟
۲. کدامیک از موارد زیر در مورد محوشدگی گرادیان (vanishing gradient) درست و کدامیک نادرست هستند؟ چرا؟
أ. تابع tanh نسبت به تابع sigmoid مشکل کمتری از لحاظ محوشدگی گرادیان ایجاد می کند.
ب. محوشدگی گرادیان باعث می شود تا یادگیری در لایه های ابتدایی شبکه نسبت به لایه های انتهایی با سرعت بیشتری انجام شود.
ت. تابع فعالیت ReLU مشکل کمتری از لحاظ محوشدگی گرادیان نسبت به sigmoid ایجاد می کند.
ث. وزندهی اولیه می تواند از بروز مشکل محوشدگی گرادیان جلوگیری کند.
۳. در یک مدل پرسپترون با تابع فعالیت sigmoid که فرایند یادگیری برای آن کامل شده است را در نظر بگیرید، نصف کردن وزن های ورودی چه تاثیری بر روی دقت عملکرد آن خواهد داشت؟
۴. یک شبکه عصبی چهار لایه را در نظر بگیرید که در ابتدای یادگیری همه وزن های آن با مقدار ۰/۵ مقداردهی اولیه شده است. آیا این روش وزندهی مناسب است؟ چرا؟
۵. اگر داده های ورودی در یک مسئله نسبت به هم کاملاً مستقل باشند، آنگاه مزایا و معایب استفاده از روش mini-batch gradient نسبت به روش stochastic gradient چیست؟
۶. برتری استفاده از بهینه گر Adam نسبت به بهینه گر stochastic gradient چیست؟ شرح دهید.
۷. دو مجموعه مجزا و هم اندازه یکی حاوی داده های کلاست مثبت و دیگری حاوی داده های کلاست منفی را در نظر بگیرید. آنگاه سناریوهای مختلف برای تلفیق این داده ها در یک مجموعه داده آموزشی که باعث شکست شبکه عصبی یادگیرنده در حل مسئله مورد نظر خواهد شد را نام برده و دلایل شکست آن ها را شرح دهید.
۸. چرا در شبکه های عصبی عمیق می بایست از توابع فعالیت غیر خطی بهره برد؟
۹. چرا استفاده از شبکه های عصبی پیچشی (convolutional) برای مجموعه داده تصاویر نسبت به شبکه های تماماً متصل (fully-connected) ارجحیت دارد؟
۱۰. انواع لایه های به کار رفته در شبکه های عصبی پیچشی را توضیح داده و فرایند پیش رو و پس انتشار خطا را برای آن ها شرح دهید.
۱۱. مسئله شناسایی تصاویر زرافه از غیر-زرافه را در نظر بگیرید به طوری که داده های جمع آوری شده برای این مسئله نامتوازن و به صورت زیر باشد:
أ. ۲۰۰ تصویر زرافه
ب. ۲۰۰۰ تصویر غیر-زرافه
الف- دو روش مناسب افزایش داده ها (data augmentation) مناسب برای این مسئله پیشنهاد دهید.
ب- اگر به جای افزایش داده ها از شبکه زیر برای حل مسئله عدم توازن داده ها استفاده شود، نقش پارامترهای α و β چیست و مقدار پیشنهادی شما برای این دو پارامتر چیست؟



$$z_1 = W_1 x^{(i)} + b_1$$

$$a_1 = \text{ReLU}(z_1)$$

$$z_2 = W_2 a_1 + b_2$$

$$\hat{y}^{(i)} = \sigma(z_2)$$

$$L^{(i)} = \alpha * y^{(i)} * \log(\hat{y}^{(i)}) + \beta * (1 - y^{(i)}) * \log(1 - \hat{y}^{(i)})$$

$$J = -\frac{1}{m} \sum_{i=1}^m L^{(i)}$$

در این شبکه z_i و a_i مقدار فعالیت-پیشین و خروجی لایه i هستند. پارامترهای W_i و b_i به ترتیب وزن‌ها و بایاس لایه i -ام هستند. $x^{(i)}$ و $y^{(i)}$ به ترتیب ورودی و خروجی مطلوب نمونه i -ام و $\hat{y}^{(i)}$ خروجی شبکه برای این نمونه هستند. همچنین $L^{(i)}$ و J به ترتیب تابع هزینه برای نمونه i -ام و تابع هزینه بر روی کل مجموعه داده هستند.

ج- برای شبکه ارائه شده در قسمت ب، مقادیر $\frac{\partial J}{\partial \hat{y}}$ ، $\frac{\partial \hat{y}}{\partial z_2}$ ، $\frac{\partial z_2}{\partial a_1}$ و $\frac{\partial a_1}{\partial w_1}$ را بدست آورید.

د- اگر یک ترم تنظیم‌گر (Regularization) از نوع L_2 به تابع هزینه اضافه گردد آن‌گاه تابع هزینه شبکه ارائه شده در قسمت ب به چه صورت خواهد شد؟

۱۲. برتری شبکه LSTM نسبت به شبکه بازگشتی عادی RNN چیست؟ مکانیزم درونی یک واحد LSTM را شرح دهید.

۱۳. وجه تمایز variational autoencoder ها به نسبت autoencoder عادی چیست؟ مکانیزم عملکرد هریک را شرح دهید.

۱۴. سوال امتیازی: برای بهتر شدن کیفیت آموزش مجازی چه راهکارهایی پیشنهاد می‌کنید؟