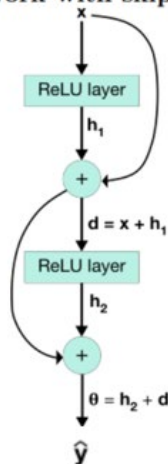


۱. (۳۰ نمره) به سوالات زیر کوتاه پاسخ دهید:

- یک شبکه عصبی fully connected را در نظر بگیرید که تابع فعالسازی تمام لایه‌ها تابع sigmoid می‌باشد. برای مقداردهی اولیه وزن‌ها، همه وزن‌های شبکه را مقادیری بزرگ انتخاب می‌کنیم. آیا این ایده خوبی است؟ استفاده از این مقداردهی اولیه موجب چه پدیده‌ای می‌شود؟
- شما در حال طراحی یک سیستم یادگیری عمیق برای تشخیص سرطان سینه با استفاده از تصاویر X-ray هستید به نظر شما مناسب‌ترین معیار ارزیابی در این مدل چه چیزی می‌تواند باشد و چرا: Accuracy, Precision, Recall, F1 score.
- شما در حال طراحی یک مدل برای یک تسک طبقه‌بندی (classification) هستید. در ابتدا مدل خود را بر روی ۲۰ نمونه آموزش می‌دهید و مشاهده می‌کنید که با وجود همگرا شدن آموزش، خطای آموزش بر روی این نمونه‌ها زیاد است. پس در ادامه تصمیم می‌گیرید که شبکه خود را این بار روی ۱۰۰۰۰ نمونه آموزش دهید. آیا روش شما برای حل این مشکل صحیح است؟ اگر بلی، محتمل‌ترین نتایج مدل خود را در این حالت توضیح دهید. اگر خیر، راه‌حلی برای رفع این مشکل بیان کنید.
- هدف استفاده از کانولوشن ۱\*۱ چیست؟

۲.

Neural network with skip connections



$$z_1 = W_1 x + b_1$$

$$h_1 = \text{ReLU}(z_1)$$

$$d = h_1 + x$$

$$z_2 = W_2 d + b_2$$

$$h_2 = \text{ReLU}(z_2)$$

$$\theta = h_2 + d$$

$$J = \text{CE}(y, \hat{y})$$

ابعاد متغیرها را به صورت  $x \in \mathbb{R}^{D_x \times 1}$ ,  $W_1 \in \mathbb{R}^{H \times D_x}$ ,  $b_1 \in \mathbb{R}^H$ ,  $W_2 \in \mathbb{R}^{D_y \times H}$ ,  $b_2 \in \mathbb{R}^{D_y}$  و  $\hat{y} \in \mathbb{R}^{D_y \times 1}$  در نظر بگیرید. همچنین فرض کنید که  $D_x = D_y = H$  است. در این سوال قصد داریم  $\frac{\partial J}{\partial x}$  را محاسبه کنیم. به ترتیب مراحل زیر را برای ایجاد جواب خود طی کنید:

الف) ابتدا  $\delta_1 = \frac{\partial J}{\partial \theta}$  را بدست آورید.

ب) حال  $\delta_2 = \frac{\partial J}{\partial z_2}$  را بدست آورید.

ج)  $\delta_3 = \frac{\partial J}{\partial d}$  را محاسبه نمایید.

د) در نهایت با توجه به نتایج بخش‌های قبل  $\frac{\partial J}{\partial \mathbf{x}}$  را بدست آورید.

۳.

(۲۰ نمره) می‌خواهیم یک شبکه CNN با معماری مشخص شده در جدول زیر را به منظور دسته‌بندی تصاویر بناهای تاریخی ایران به ۶ دسته‌ی معماری دوره‌ی ایلام (عیلام)، معماری دوره مادها، معماری دوره هخامنشیان، معماری دوره اشکانیان، معماری دوره ساسانیان و معماری اسلامی، ایجاد کنیم.

الف) برای هر لایه از شبکه با توجه به اطلاعات زیر، تعداد وزن‌ها، تعداد بایاس و سایز Feature map را در جدول وارد کنید.

- منظور از CONV-K-N یک لایه ی کانولوشنی با N فیلتر با سایز K\*K است. Stride و Padding برای همه ی لایه های CONV به ترتیب ۰ و ۱ در نظر گرفته شده است.
- منظور از Pool-K یک K\*K pooling layer است که Stride برابر با K و Padding برابر با صفر دارد.
- منظور از FC-N یک لایه ی Fully Connected با N نورون است.

Layer	Activation Map Dimension	Number of Weights	Number of Biases
Input	۱۲۸*۱۲۸*۳	.	.
CONV-9-32			
Pool-2			
CONV-5-64			
Pool-2			
CONV-5-64			
Pool-2			
FC-6	۶	۶(۱۲*۱۲*۶۴)	۶

ب) چه تابع فعال سازی برای لایه آخر استفاده می کنید؟ با ذکر فرمول بیان کنید چه مزیتی نسبت به سایر توابع دارد.

ج) آیا می‌توانیم برای بهبود یادگیری الگوهای پیچیده تر، لایه های بیشتری به شبکه اضافه کنیم؟ با این کار چه مشکلی ممکن است برای شبکه ایجاد شود؟

د) دلیل استفاده از لایه های Pooling چیست؟

ه) استفاده از Stride بزرگتر از ۱ چه فایده ای دارد؟

(۲۰ نمره) تابع فعالسازی ReLU می‌تواند باعث بوجود آمدن نورون‌های مرده، یعنی نورون‌هایی که تحت هیچ ورودی فعال نمی‌شوند و خروجی آن‌ها به ازای هر ورودی صفر است شود. شبکه عصبی feedforward دو لایه را با  $N$  نورون ورودی و  $H$  نورون در لایه پنهان (hidden) در نظر بگیرید که وزن میان آن‌ها  $\mathbf{W}^{(1)}$  و بایاس آن‌ها نیز  $b^{(1)}$  می‌باشد. خروجی این شبکه نیز یک نورون داشته (خروجی اسکالر) و با وزن‌های  $W^{(2)}$  است.

$$h_i = \text{ReLU}(W_i^{(1)} \cdot x + b_i^{(1)}) = \text{ReLU}\left(\sum_{j=1}^N W_{ij}^{(1)} x_j + b_i^{(1)}\right) \quad \text{for } i \in \{1, 2, \dots, H\}$$

$$\hat{y} = W^{(2)} \cdot h$$

که در آن هدف ما بهینه کردن تابع هزینه‌ی مشتق‌پذیر دلخواه  $l : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  که آرگومان‌های آن برجسب واقعی داده‌ها و پیش‌بینی‌های شبکه است، می‌باشد.

الف) تحت چه شرایطی نورون  $h_i$  مرده است؟ جواب شما باید بر حسب  $W_i^{(1)}$  یعنی سطری از  $\mathbf{W}^{(1)}$  که متناظر نورون  $i$  می‌باشد، و  $b^{(1)}$  بیان شود.

ب) فرض کنید که به ازای یک نمونه خواهیم داشت:  $\frac{\partial l}{\partial y} = 1$ . مشتقات جزئی  $\frac{\partial l}{\partial W_{ij}^{(1)}}$  و  $\frac{\partial l}{\partial b_i^{(1)}}$  را برای این نمونه بدست آورید.

ج) با توجه به نتایج خود در بخش‌های قبلی، توضیح دهید چرا یک نورون مرده نمی‌تواند زنده شود؟! منظور از زنده شدن این است که پارامترهای آن به گونه‌ای تغییر کنند که به ازای تمام ورودی‌ها خروجی این نورون صفر نباشد.

د) راه‌حلی برای رفع این مشکل تابع ReLU ارائه دهید. (لزوماً جواب یکتایی برای این بخش وجود ندارد، اما باید راه‌حل پیشنهادی خود را توجیه کنید.)