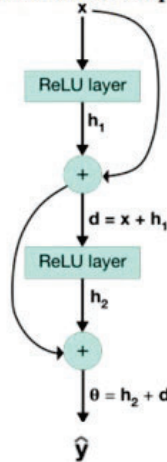


۱. (۳۰ نمره) به سوالات زیر کوتاه پاسخ دهید:

- 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.