

(سریع و آسان)

الف) تابع  $sgn$  مشکل ~~vanishing gradient~~ دارد

ب) تابع  $y=x$  ~~vanishing gradient~~ ندارد  
کتابخانه خیلی است و عددی کاری نمی کند یعنی وقتی آن را در شبکه به اعم افکار حل شبکه یک کسری است  
تا اینجا خدو بی آن در یک  $range$  نیست و ~~exploding gradient~~ است

تابع  $y=sgn x$  به هم خدو بی آن در یک  $range$  نیست و مشکل است ~~exploding gradient~~  
مشکل ~~vanishing gradient~~ دارد

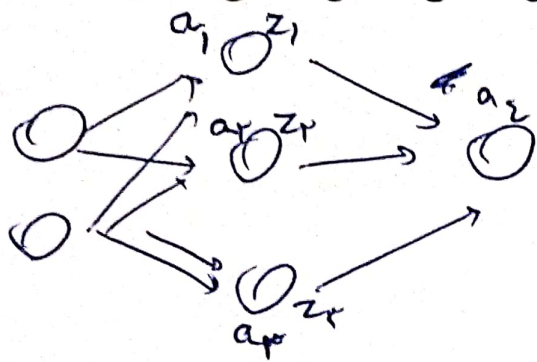
ب) در واقع به دنبال این است که گزاردان به جلات قبلی هم وابسته باشند به همین علت یک محاسبه در کتاب جلات قبل به گزاردان اضافه می کند که باعث می شود محاسبه سریع تر انجام شود.

ج) Relu  
✓ پیوسته است  $x$  محدود نیست  
✓ محاسبه ساده است  
✓ مشتق آن به آسانی  
 $x > 0$  غیر صفر است  
~~vanishing~~  $x$  ~~محاسبه~~ مشکل  
~~dead~~  $x$  ~~محاسبه~~ مشکل  
neuron  
tanh  
✓ پیوسته  
✓ مشتق پیوسته  
✓ محدود  
✓ محاسبه مشتق آسان  
✓ محاسبه مشتق آسان  
✓ مشتق اصلی  
vanishing  $x$  ~~gradient~~

(۲) الف چون در واقع ~~این~~ در نقطه کمین مقدار بیش از  $\text{variance}$  (واریانس) کمی شود و پستی می شود

سیر پستی تا نقطه بهینه را هلی کنیم (در واقع تأثیر تولید کمتر شود)

(ب) وقتی اثر از به قدر کافی بزرگ است واریانس در نتیجه مقدار زیاد تأثیر کم  
دستبرد و وقتی از تابع را smooth می کند در واقع به کمک ~~این~~ واریانس ثابت خواهد بود



$$z_1 a_1 = x_1 w_1 + x_2 w_2 = 4$$

$$z_1 = \text{ReLu}(a_1) = 4$$

$$a_2 = x_1 w_3 + x_2 w_4 = 1 + (-2) = -1$$

$$z_2 = \text{ReLu}(a_2) = 0$$

$$a_3 = x_1 w_5 + x_2 w_6 = 1 + 2 = 3$$

$$z_3 = \text{ReLu}(a_3) = 3$$

$$a_4 = w_7 z_1 + w_8 z_2 + w_9 z_3 = -2 + 0 + 1.2 \times 3 = 1.6$$

$$\text{ReLu}(a_4) = 1.6$$

$$L = \text{Loss} = \frac{1}{2} [y - \hat{y}]^2 = \frac{1}{2} (6 - 3)^2 = \frac{9}{2}$$

$$\frac{\partial L}{\partial w_9} = \frac{\partial L}{\partial a_4} \times \frac{\partial a_4}{\partial w_9} = \frac{3}{4} \times 5 = 22.5$$

$$\frac{\partial L}{\partial w_8} = \frac{\partial L}{\partial a_4} \times \frac{\partial a_4}{\partial w_8} = \frac{3}{4} \times 0 = 0$$

$$\frac{\partial L}{\partial w_7} = \frac{\partial L}{\partial a_4} \times \frac{\partial a_4}{\partial w_7} = \frac{3}{4} \times 1.2 = 0.9$$



$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial a_4} \frac{\partial a_4}{\partial z_1} \frac{\partial z_1}{\partial a_1} \frac{\partial a_1}{\partial w_1} = 15$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial a_4} \frac{\partial a_4}{\partial z_1} \frac{\partial z_1}{\partial a_1} \frac{\partial a_1}{\partial w_2} = -6$$

$$\frac{\partial L}{\partial w_3} = \frac{\partial L}{\partial a_4} \frac{\partial a_4}{\partial z_2} \frac{\partial z_2}{\partial a_2} \frac{\partial a_2}{\partial w_3} = 10$$

$$\frac{\partial L}{\partial w_4} = \frac{\partial L}{\partial a_4} \frac{\partial a_4}{\partial z_2} \frac{\partial z_2}{\partial a_2} \frac{\partial a_2}{\partial w_4} = 0$$

$$\frac{\partial L}{\partial w_5} = \frac{\partial L}{\partial a_4} \frac{\partial a_4}{\partial z_3} \frac{\partial z_3}{\partial a_3} \frac{\partial a_3}{\partial w_5} = 6$$

$$\frac{\partial L}{\partial w_6} = \frac{\partial L}{\partial a_4} \frac{\partial a_4}{\partial z_3} \frac{\partial z_3}{\partial a_3} \frac{\partial a_3}{\partial w_6} = 12$$

$$w_1 = 2 - 1 \times 12.50$$

$$w_2 = 1 - 1 \times 12.50$$

$$w_3 = w_3$$

$$w_4 = w_4$$

$$w_5 = 1 - 1 \times 12.50$$

$$w_6 = 2 - 1 \times 12$$

$$w_7 = 1 - 1 \times 12$$

$$w_8 = w_8$$

$$w_9 = 1 - 1 \times 12$$