$\cancel{f \Rightarrow f'}$

1. $f(x) = kx, \ x \in R$

$\Rightarrow f'(x) = k$

2. $\cancel{f(x) = x}$

$f(x,y) = f_1(x) f_2(y)$

$\Rightarrow \frac{\partial f}{\partial x} = f_2(y) \cdot f_1'(x)$

3. $f(x) = f_1(x) f_2(x)$

$\Rightarrow f'(x) = f_1'(x) f_2(x) + f_2'(x) f_1(x)$

$\cancel{4. \ f'(x) = Ax, \ x \in R^n, \ A \in R^{m \times n}}$

4. $f(x) = a^T x, \quad x \in R^n, \ a \in R^n$

$\Rightarrow f'(x) = a^T,$

5. $f(x) = Ax, \quad x \in R^n, \ A \in R^{m \times n}$

$\Rightarrow f'(x) = A$

$\cancel{6. \ f(x,y) = f_1(x)^T f_2(y), \ f}$

6. $f(x) = (f_1 \cdot f_2)(x), \quad f_2(x) \in R, x \in R, f_1 \in R$

$\cancel{\Rightarrow f'(x) = (f_1' \cdot f_2')(x)}$

solve: $f(x) = f_1(f_2(x))$

$\therefore f'(x) = f_1'(f_2(x)) \cdot f_2'(x)$

7. $\cancel{f(x,y) = f_1(x)^T}$

$f(x) = a^T f_1(x),$

$x \in R^n, f_1 \in R^n$

$\Rightarrow f'(x) = a^T f_1'(x)$

8. $f(x,y) = \cancel{f_1(x)^T \cdot f_2(y)}$

$x \in R^n, \ f_1(x) \in R^m,$

$f_2(y) \in R^m$

$\Rightarrow f_x'(x,y) = f_2(y)^T \cdot f_1'(x)$

9. $f(x) = f_1(x)^T \cdot f_2(x)$

$f_1(x) \in R^m, \ f_2(x) \in R^m,$

$x \in R^n$

$\Rightarrow f'(x) = f_1(x)^T \cdot f_2'(y)$

$+ f_2(y)^T f_1'(x)$

10: $f(x) = f_2(f_1(x))$

$x \in R^n, \ f_1(x) \in R^m,$

$f_2(\cdot) \in R$

$\Rightarrow f'(x) = f_2'(f_1(x)) \cdot f_1'(x)$

11. $f(x) = (f_2 \cdot f_1)(x),$

$f_1(x) \in R^m, \ f_2(x) \in R^{m_2}$

$x \in R^n$

$\Rightarrow f'(x) = f_2'(f_1(x)) \cdot f_1'(x)$

$= f_2' \cdot f_1'$

12. $f(x) = Wx + b$

$\Rightarrow f'(x) = W$

life ~~hobby~~ ~~strength~~ movie mountain time read not ~~ok~~

13. $f(x, W, b) = Wx + b \in R^m,$

$\Rightarrow \frac{\partial f}{\partial W} = \begin{pmatrix} x^T & & \\ & x^T & \\ & & x^T \end{pmatrix} \in R^{m \times (m \cdot n)}$

$\frac{\partial f}{\partial b} = I \in R^{m \times m}$

$\frac{\partial f}{\partial x} = W$

$\cancel{14. \text{ logistic function: } f(x) = \frac{L}{1+e^{-k(x-x_0)}}}$

14. $y = f(x) = \frac{1}{1+e^{-x}} \Rightarrow f' = y(1-y)$

15. logistic function:

$z = f(t) = \frac{L}{1+e^{-k(t-t_0)}}$

$\Rightarrow f' = L \cdot (y(1-y)) \cdot k$

$= \frac{kz}{L}(L - z)$

here $y = \frac{1}{1+e^{-x}}, \quad x = k(t - t_0)$

16. relu: $y = f(x) = x \cdot (x>0)$

$\Rightarrow f'(x) = (y>0)$

17. tanh:

$y = f(x) = \dfrac{e^{2x}-1}{e^{2x}+1}$

$\Rightarrow f'(x) = 1 - y^2$

18. $y = f(x)$, $x \in R^n$

$y_j = \dfrac{e^{x_j}}{\sum_k e^{x_k}}$

softmax function.

$\Rightarrow \dfrac{\partial y_i}{\partial x_j} = \begin{cases} y_i(1-y_j), & i=j \\ -y_i y_j, & i \neq j \end{cases}$

$= y_i(1_{i=j} - y_j)$

$\therefore f'(x) = diag(y) - y \cdot y^T$

19. $f(x) = -\sum a_i \log x_i$

$a_i = \begin{cases} 1, & i=r \\ 0, & else \end{cases}$

$\Rightarrow f(x) = -\log x_r$

$f' = g(x) = (\cdots, 0, -\frac{1}{x_r}, 0, \cdots)$

here $g_i = \begin{cases} -\frac{1}{x_i}, & i=r \\ 0, & else. \end{cases}$

20. $z = f(x) = (f_1 f_2)(x)$

$z = f_1(y) = -\sum a_i \log y_i$,

$a_i = \begin{cases} 1, & i=r \\ 0, & else \end{cases}$

$y = f_2(x) = \frac{1}{\|e^x\|_1} e^x$,

$x \in R^n$

$\Rightarrow f'(x) = g(x) = (\cdots, y_{r-1}, y_r - 1, y_{r+1}, \cdots)$

i.e., $g_i = \begin{cases} y_i - 1, & i=r \\ y_i, & else \end{cases}$

proof:

$g = (\cdots, 0, -\frac{1}{y_r}, 0, \cdots) \cdot (diag(y) - yy^T)$

$= -\frac{1}{y_r}(-y_r y_1, -y_r y_2, \cdots, y_r - y_r^2, \cdots)$

$= (y_1, y_2, \cdots, y_{r-1}, y_r - 1, y_{r+1}, \cdots)$

21. $z = f(y)$, $y = Wx + b$,

$x \in R^n$, $y \in R^m$, $z \in R$.

$\Rightarrow \dfrac{\partial z}{\partial W} = f'(y)^T \cdot x^T$, $\dfrac{\partial z}{\partial b} = f'(y)$, $\dfrac{\partial z}{\partial x} = f'(y) \cdot W$.

here $f'(y)$ is row vector.

proof: $\dfrac{\partial z}{\partial \tilde{W}} = f'(y)(x^T \; x^T \; \cdots \; x^T)$

~~here~~ $\tilde{W}$ is vector, ~~reshaped W~~.

$\tilde{W}$ vector $\leftrightarrow$ W matrix.

$\therefore \dfrac{\partial z}{\partial W} = f'(y)^T x^T$.

$\therefore$

21. $z = f(Y)$, $Y = WX + b$

$z \in R$, $Y \in R^{m\times b}$, $X \in R^{n\times b}$

$$\Rightarrow \frac{\partial z}{\partial W} = f'(Y)X^T$$

prove:

① method 1:

$$\because \frac{\partial z}{\partial W_{ij}} = \sum_{k=1}^{b} \frac{\partial z}{\partial y_i^{(k)}} x_j^{(k)}$$

$$\therefore \frac{\partial z}{\partial W} = \frac{\partial z}{\partial Y} \cdot X^T$$

② method 2:

let $Y = \begin{pmatrix} Y_1^T \\ \vdots \\ Y_m^T \end{pmatrix}$

$$\because \frac{\partial z}{\partial \tilde{W}} = \frac{\partial z}{\partial \tilde{Y}} \cdot \frac{\partial \tilde{Y}}{\partial \tilde{W}} = \left(\frac{\partial z}{\partial Y_1}, \frac{\partial z}{\partial Y_m}\right) \cdot \begin{pmatrix} X^T & & \\ & \ddots & \\ & & X^T \end{pmatrix}$$

$$\therefore \frac{\partial z}{\partial W} = \frac{\partial z}{\partial Y} X^T$$

proof:

$Y = (Y_1, Y_2, Y_b)$,

$X = (X_1, X_2, X_b)$.

$\because Y_i = WX_i + b$

$$\frac{\partial Y_i}{\partial \tilde{W}} = \begin{pmatrix} X_i^T & & \\ & & \\ & & X_i^T \end{pmatrix}$$

$$\therefore \frac{\partial \tilde{Y}}{\partial \tilde{W}} = \begin{pmatrix} X^T & & \\ & \ddots & \\ & & X^T \end{pmatrix}$$

26. $z = f(y)$, $y = Wx + b$,

$x \in R^n$, $y \in R^m$, $z \in R$

$$\Rightarrow \frac{\partial z}{\partial \tilde{W}} = \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial \tilde{W}} = \left(\frac{\partial z}{\partial y_1} \frac{\partial z}{\partial y_m}\right) \cdot \begin{pmatrix} x^T & \\ & x^T \end{pmatrix}$$

$$\therefore \frac{\partial z}{\partial W} = f'(y)^T \cdot x^T$$

22. $z = f(x, y)$,

$x = f_1(t)$, $y = f_2(t)$

$t \in R^n$, $z \in R$, $x \in R^{m_1}$, $y \in R^{m_2}$

$$\Rightarrow \frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial t}$$

23. $z = f(y)$, $y = \tanh x$,

$z \in R$, $x, y \in R^n$

$$\Rightarrow \frac{\partial z}{\partial x} = \frac{\partial z}{\partial y}(I - \mathrm{diag}(y \odot y)) = \frac{\partial z}{\partial y} - \frac{\partial z}{\partial y} \odot y \odot y$$

24. $y = Wx + b$, $x \in R^n$, $y \in R^n$

$$\Rightarrow \frac{\partial y}{\partial \tilde{W}} = \begin{pmatrix} x^T & & \\ & x^T & \\ & & x^T \end{pmatrix} \in R^{m\times(m\times n)}$$

25. $Y = WX + b$. $X \in R^{n\times b}$; $Y \in R^{m\times b}$

$$\Rightarrow \frac{\partial \tilde{Y}}{\partial \tilde{W}} = \begin{pmatrix} X^T & & \\ & X^T & \\ & & X^T \end{pmatrix} \in R^{(m\times b)\times(m\times n)}$$

28. $y = Wx + b$, $x \in R$, $y \in R$

$\Rightarrow \frac{\partial y}{\partial x} = W$, $\frac{\partial y}{\partial b} = I$

29. $z = f(y)$, $y = Wx + b$

$x \in R^n$, $y \in R^m$, $z \in R$

$\Rightarrow \frac{\partial z}{\partial x} = \frac{\partial z}{\partial y} W$.

$\frac{\partial z}{\partial b} = \frac{\partial z}{\partial y}$

30. $y = w^T x + b$, $x \in R^n$, $y \in R$

$\Rightarrow \frac{\partial y}{\partial x} = w^T$, $\frac{\partial y}{\partial b} = 1$

31. $y = w^T X + b$,

$X \in R^{n \times b}$, $y \in R^b$, $b \in R$.

$\Rightarrow \frac{\partial y}{\partial \tilde{X}} = \begin{pmatrix} vec(w, 0, 0) \\ vec(0, w, 0) \\ vec(0, 0, w) \end{pmatrix}$

solve: $\because y_i = w^T \cancel{X_i}$, $\frac{\partial y_i}{\partial X_i} = w^T$

here $y = (y_1, y_b)$, $X = (X_1, X_b)$

$\frac{\partial y_i}{\partial X} = (0, \underset{i}{w}, 0)$, $\forall i \in \{1, \cdots, b\}$

$\therefore \frac{\partial y}{\partial \cancel{X}} = \begin{pmatrix} \frac{\partial y_1}{\partial \cancel{X}} \\ \frac{\partial y_b}{\partial \cancel{X}} \end{pmatrix}$

I	You
Mom.	movie
math.	book
work	
eng?	Eng?

$author^{?} = \begin{pmatrix} vec(w, 0, 0) \\ vec(0, w, 0) \\ vec(0, 0, w) \end{pmatrix} \in R^{b \times (n \cdot b)}$

饭

若No, say it.

Don't worry.

生命

32. $Y = WX + b$,

$X \in R^{n \times b}$, $Y \in R^{m \times b}$,

$b \in R^m$

$\Rightarrow \frac{\partial \tilde{Y}}{\partial \tilde{X}}$, $\frac{\partial \hat{Y}}{\partial \hat{X}}$, $\frac{\partial \hat{Y}}{\partial b}$

solve①: $Y = \begin{pmatrix} Y_1^T \\ Y_m^T \end{pmatrix}$, $W = \begin{pmatrix} W_1^T \\ W_m^T \end{pmatrix}$,

$b = \begin{pmatrix} b_1 \\ b_m \end{pmatrix}$.

$\because Y_i^T = W_i^T X + b_i$

$\therefore \frac{\partial Y_i}{\partial \tilde{X}} = \begin{pmatrix} vec(W_i & & \\ vec(& W_i & \\ vec(& & W_i \end{pmatrix} \in R^{b \times (n \cdot b)}$

$\therefore \frac{\partial \tilde{Y}}{\partial \tilde{X}} = \begin{pmatrix} \frac{\partial Y_1}{\partial \tilde{X}} \\ \frac{\partial Y_m}{\partial \tilde{X}} \end{pmatrix} = \begin{pmatrix} vec: W_1 & & \\ vec: & W_1 & \\ vec: & & W_1 \\ vec: W_m & & \\ vec: & W_m & \\ vec: & & W_m \end{pmatrix} \in R^{(n \cdot b) \times (n \cdot b)}$

solve②:

let $Y = (Y_1, Y_b)$, $X = (X_1, X_b)$,

$\because Y_j = WX_j + b$

$\therefore \frac{\partial Y_j}{\partial \hat{X}} = (0, \underset{j}{W}, 0)$

$\therefore \frac{\partial \hat{Y}}{\partial \hat{X}} = \begin{pmatrix} \frac{\partial Y_1}{\partial \hat{X}} \\ \frac{\partial Y_b}{\partial \hat{X}} \end{pmatrix} = \begin{pmatrix} W & & \\ & W & \\ & & W \end{pmatrix}$

here ~ mean row major vector, ^ : col major vector.

33. $z = f(Y)$,

$Y = WX + b$,

$Y \in R^{m \times b}$, $X \in R^{n \times b}$, $z \in R$

$\Rightarrow \frac{\partial z}{\partial X}$, $\frac{\partial z}{\partial b}$

solve①: $\because Y^T = X^T W^T + b^T$

$\therefore \frac{\partial z}{\partial X^T} = \frac{\partial z}{\partial Y^T} \cdot W$ (see 27)

$\therefore \frac{\partial z}{\partial X} = W^T \cdot \frac{\partial z}{\partial Y} = W^T f'(Y)$

② $\frac{\partial z}{\partial b} = \frac{\partial z}{\partial \hat{Y}} \cdot \frac{\partial \hat{Y}}{\partial b}$

$= \left(\frac{\partial z}{\partial Y_1}, \frac{\partial z}{\partial Y_b}\right) \begin{pmatrix} I \\ I \end{pmatrix}$

$= \sum_{i=1}^{b} \frac{\partial z}{\partial Y_i}$, here $Y = (Y_1, Y_b)$

solve③: $\because Y = WX + b$

$\therefore Y^T = X^T W^T + b^T$

$\therefore \frac{\partial \tilde{Y}^T}{\partial \tilde{X}^T} = \begin{pmatrix} W & & \\ & W & \\ & & W \end{pmatrix}$ (see 25)

$\therefore \frac{\partial \hat{Y}}{\partial \hat{X}} = \begin{pmatrix} W & & \\ & W & \\ & & W \end{pmatrix}$

④ $\frac{\partial \hat{Y}}{\partial b} = \begin{pmatrix} I \\ I \\ I \end{pmatrix} \in R^{(m \times b) \times m}$

34. RNN function:

$h^{(t)} = \tanh(W_{hh} h^{(t-1)} + W_{xh} x^{(t)} + b_h)$

$y^{(t)} = W_{hy} h^{(t)} + b_y$,

$p^{(t)} = \mathrm{softmax}(y^{(t)})$

$L = -\sum_{t=1}^{T} r^{(t)} \log p^{(t)}$

$r^{(t)}$ is one-hot, i.e., $r^{(t)}(k) = \begin{cases} 1, & k=r \\ 0, & \text{else} \end{cases}$

$\Rightarrow \frac{\partial L}{\partial W_{hy}}, \frac{\partial L}{\partial b_y}, \frac{\partial L}{\partial W_{hh}}, \frac{\partial L}{\partial W_{xh}}, \frac{\partial L}{\partial b_h}$.

solve: $\forall t = T, \cdots, 1$.

① $\frac{\partial L}{\partial y^{(t)}} = \frac{\partial L}{\partial p^{(t)}} \frac{\partial p^{(t)}}{\partial y^{(t)}} = (p_{r-1}, \underset{r}{p_r - 1}, p_{r+1})$

(see 20)

② $\frac{\partial L}{\partial W_{hy}} \mathrel{+}= \frac{\partial L}{\partial y^{(t)}} \cdot \frac{\partial y^{(t)}}{\partial W_{hy}} = \left(\frac{\partial L}{\partial y^{(t)}}\right)^T \cdot h^{(t)T}$

Todo: develop tensor theory.

$\frac{\partial L}{\partial b_y} \mathrel{+}= \frac{\partial L}{\partial y^{(t)}}$

③ $\frac{\partial L}{\partial h^{(t)}} = \frac{\partial L}{\partial y^{(t)}} \frac{\partial y^{(t)}}{\partial h^{(t)}} + \left(\frac{\partial L}{\partial h^{(t)}} \text{ at } t+1\right)$

$= \frac{\partial L}{\partial y^{(t)}} W_{hy} + \left(\frac{\partial L}{\partial h^{(t)}} \text{ at } t+1\right)$

④ $\frac{\partial L}{\partial h^{(t)}_{raw}} = \frac{\partial L}{\partial h^{(t)}} \cdot \frac{\partial h^{(t)}}{\partial h^{(t)}_{raw}}$

$= \frac{\partial L}{\partial h^{(t)}} \odot (1 - h^{(t)} \odot h^{(t)})$

(see 23)

⑤ $\frac{\partial L}{\partial W_{hh}} \mathrel{+}= \left(\frac{\partial L}{\partial h^{(t)}_{raw}}\right)^T (h^{(t-1)})^T$

$\frac{\partial L}{\partial W_{xh}} \mathrel{+}= \left(\frac{\partial L}{\partial h^{(t)}_{raw}}\right)^T x^{(t)T}$

$\frac{\partial L}{\partial b_h} \mathrel{+}= \frac{\partial L}{\partial h^{(t)}_{raw}}$

⑥ $\left(\frac{\partial L}{\partial h^{(t-1)}} \text{ at } t\right) = \frac{\partial L}{\partial h^{(t)}_{raw}} W_{hh}$

35. mlp layer:

$Z = f(Y)$, $Y = WX + b$,

$L = L(Z)$, f is element-wise.

$L \in R$, $Z, Y \in R^{m \times b}$, $X \in R^{n \times b}$

$\Rightarrow \frac{\partial L}{\partial W}, \frac{\partial L}{\partial b}, \frac{\partial L}{\partial X}$.

solve: $\frac{\partial L}{\partial Y} = \frac{\partial L}{\partial Z} \odot f'(Y)$

here f' is element-wise.

$\therefore \frac{\partial L}{\partial W} = \frac{\partial L}{\partial Y} \cdot X^T$

$\frac{\partial L}{\partial b} = \sum_{i=1}^{b} \frac{\partial L}{\partial Y_i}$, $Y = (Y_1, Y_i, Y_b)$

$\frac{\partial L}{\partial X} = W^T \frac{\partial L}{\partial Y}$

36. batch normalization function:

$$Y=f(X),\ X\in R^{n\times m}$$

detail:

$$\mu=\frac{1}{m}\Sigma X_i,\quad X=(X_1, X_i, X_m)$$

$$\sigma^2=\frac{1}{m}\Sigma(X_i-\mu)^2,\quad \mu\in R^n,\ \sigma^2\in R^n$$

$$\hat{X}_i=\frac{X_i-\mu}{\sqrt{\sigma^2+\varepsilon}},\quad \varepsilon=10^{-6}$$

$$Y_i=k\hat{X}_i+b,\quad Y=(Y_1, Y_i, Y_m)$$

$\frac{\partial l}{\partial Y}$ is known, here $l=l(Y)\in R$, $k\in R^n$, $b\in R^n$

$$\Rightarrow \frac{\partial l}{\partial X},\ \frac{\partial l}{\partial k},\ \frac{\partial l}{\partial b}$$

① $\frac{\partial l}{\partial \hat{X}_i}=\frac{\partial l}{\partial Y_i}\cdot k$

② $\because \frac{\partial \hat{X}_i}{\partial \sigma^2}=-\frac{1}{2}(X_i-\mu)(\sigma^2+\varepsilon)^{-\frac{3}{2}}$

$\therefore \frac{\partial l}{\partial \sigma^2}=\sum_i \frac{\partial l}{\partial \hat{X}_i}\frac{\partial \hat{X}_i}{\partial \sigma^2}$

(此处 $\frac{\partial \hat{X}_i}{\partial \sigma^2}$ is element-wise)

③ $\because \frac{\partial \hat{X}_i}{\partial \mu}=-\frac{1}{\sqrt{\sigma^2+\varepsilon}}$

$$\frac{\partial \sigma^2}{\partial \mu}=-\frac{2}{m}\Sigma(X_i-\mu)$$

$\therefore \frac{\partial l}{\partial \mu}=\sum_i \frac{\partial l}{\partial \hat{X}_i}\frac{\partial \hat{X}_i}{\partial \mu}+\frac{\partial l}{\partial \sigma^2}\frac{\partial \sigma^2}{\partial \mu}$ solve;

④ $\because \frac{\partial \mu}{\partial X_i}=\frac{1}{m}$, (vector)

(i.e., $(\frac{1}{m},\frac{1}{m},\cdots)$)

注：·此处矩阵乘及向量内积都用·表示，具体含义根据场景判断。
求导也如此，意义不唯一，必须在场景下理解才唯一。

$$\frac{\partial \sigma^2}{\partial X_i}=\frac{2}{m}(X_i-\mu)$$

$$\frac{\partial \hat{X}_i}{\partial X_i}=\frac{1}{\sqrt{\sigma^2+\varepsilon}}$$

$\therefore \frac{\partial l}{\partial X_i}=\frac{\partial l}{\partial \hat{X}_i}\frac{\partial \hat{X}_i}{\partial X_i}+\frac{\partial l}{\partial \sigma^2}\frac{\partial \sigma^2}{\partial X_i}+\frac{\partial l}{\partial \mu}\frac{\partial \mu}{\partial X_i}$

⑤ $\frac{\partial l}{\partial k}=\Sigma\frac{\partial l}{\partial Y_i}\hat{X}_i$

$$\frac{\partial l}{\partial b}=\Sigma\frac{\partial l}{\partial Y_i}$$

37. $X, Y \in R^{d_1\times d_2\times d_3}$

$\Rightarrow \langle X,Y\rangle := \sum_{i,j,k} X(i,j,k)\cdot Y(i,j,k)$

is inner product.

38. $Wx = \begin{pmatrix}\langle w_1,x\rangle\\ \vdots \\ \langle w_m, x\rangle\end{pmatrix}$

here $W=\begin{pmatrix}w_1^T\\ \vdots\\ w_m^T\end{pmatrix}\in R^{m\times n}$, $x\in R^n$

39. $Y=WX$, $W=\begin{pmatrix}w_1^T\\ w_m^T\end{pmatrix}\in R^{m\times n}$

$X=(x_1,\cdots,x_b)\in R^{n\times b}$, $Y\in R^{m\times b}$

$\Rightarrow Y(i,j)=\langle w_i, x_j\rangle$,

(矩阵乘)

40. $Y=WX$,

$W\in R^{(m_1\times m_2)\times(n_1\times n_2)}$

$X\in R^{(n_1\times n_2)\times(b_1\times b_2)}$

$Y\in R^{(m_1\times m_2)\times(b_1\times b_2)}$

$\Rightarrow Y(i_1,i_2,j_1,j_2)$

$:= Y((i_1,i_2),(j_1,j_2))$

$=\langle W(i_1,i_2,:,:), X(:,:,j_1,j_2)\rangle$

也是矩阵乘. (更广义的)

41. 1D卷积

$Y=f(X)=K\odot X$, $K\in R^{n_k}$, $X\in R^n$

$Y\in R^m$

$\Rightarrow Y(i)=\langle K, X'\rangle$, $X'=X(i:i+n_k)$

41.3 $Y=f(X)=K\odot X$, $K\in R^{n_k}$, $X\in R^n$, $Y\in R^m$

given $\frac{\partial L}{\partial Y}$, here $L=L(Y)\in R$,

$\Rightarrow \frac{\partial L}{\partial K}$, $\frac{\partial L}{\partial X}$.

solve: ① define $X'\in R^{n_k\times m}$,

$m=n-n_k+1$.

let $X'=(X'_1,\cdots,X'_m)$

$X'_i := X(i:i+n_k) = (X(i),\cdots,X(i+n_k-1))$

则 $Y=KX'$ (Y是向量,行向量或列向量)

(将卷积转化成矩阵乘)

② ∴ $\frac{\partial L}{\partial K}=\frac{\partial L}{\partial Y}X'^T$ (见27题)

(这里$\frac{\partial L}{\partial Y}$是行向量)

$\frac{\partial L}{\partial X'}=K^T\frac{\partial L}{\partial Y}$ (见33题)

③ ∵ 函数关系 $X\to X'$

有 $X'(j,i)=X(i+j-1)$

∴ 我们如下求$\frac{\partial L}{\partial X}$: 先令$\frac{\partial L}{\partial X}=0\in R^n$

$\frac{\partial L}{\partial X(i+j-1)} \leftarrow \frac{\partial L}{\partial X(i+j-1)}+\frac{\partial L}{\partial X'(j,i)}$

$\forall i,j$

∴ $\frac{\partial L}{\partial X}$便求得.

42. 2D卷积

$Y=f(X)=K\odot X$,

$K\in R^{n_{k_1}\times n_{k_2}}$, $X\in R^{n_1\times n_2}$, $Y\in R^{m_1\times m_2}$

$\Rightarrow Y(i_1,i_2) = \langle K, X' \rangle,$

$X' = X(i_1 : i_1+n_{k_1},\ i_2 : i_2+n_{k_2})$

42.3 已知 $Y=f(X)=K\odot X$, (条件如42)

given $\frac{\partial L}{\partial Y}$, $L=L(Y)\in R$,

$\Rightarrow \frac{\partial L}{\partial K}, \frac{\partial L}{\partial X}$.

solve:

① define $X' \in R^{(n_{k_1}\times n_{k_2})\times(m_1\times m_2)}$

$X'(j_1,j_2,i_1,i_2)$

$= X(i_1+j_1-1,\ i_2+j_2-1)$

here $m_1 = n_1 - n_{k_1}+1$,

$m_2 = n_2 - n_{k_2}+1$.

$\therefore Y = KX'$

(广义上的向量K与矩阵X'乘法，定义见40题.)

② $\therefore \frac{\partial L}{\partial K} = \frac{\partial L}{\partial Y} X'^T$ (27题)

$\frac{\partial L}{\partial X'} = K^T \frac{\partial L}{\partial Y}$ (33题)

③ 令 $\frac{\partial L}{\partial X} = 0 \in R^{n_1\times n_2}$

$\forall\ j_1\in[1,n_{k_1}],\ j_2\in[1,n_{k_2}]$

$i_1\in[1,m_1],\ i_2\in[1,m_2]$,

$\frac{\partial L}{\partial X(i_1+j_1-1,\ i_2+j_2-1)} \mathrel{+}= \frac{\partial L}{\partial X'(j_1,j_2,i_1,i_2)}$

$\therefore$ we get $\frac{\partial L}{\partial X}$.

43. fully connected function:

$Y=f(X)=WX$,

$W\in R^{(m_1\times m_2)\times(n_1\times n_2\times n_3)}$

$X\in R^{n_1\times n_2\times n_3}$, $Y\in R^{m_1\times m_2}$

43.3 if $W\in R^{(m_1\times m_2\times m_3)\times(n_1\times n_2\times n_3)}$

$\Rightarrow Y\in R^{m_1\times m_2\times m_3}$

43.4. 已知 $Y=f(X)$ 为43中的函数，

$W\in R^{(m_1\times m_2\times m_3)\times(n_1\times n_2\times n_3)}$

$X\in R^{n_1\times n_2\times n_3}$, $Y\in R^{m_1\times m_2\times m_3}$,

given $\frac{\partial L}{\partial Y}$, here $L=L(Y)\in R$,

$\Rightarrow \frac{\partial L}{\partial W}, \frac{\partial L}{\partial X}$.

solve:

① $\because Y(i_1,i_2,i_3) = \langle W(i_1,i_2,i_3,:,:,:), X\rangle$

$\therefore \frac{\partial L}{\partial W} = \frac{\partial L}{\partial Y} X^T$

② $\because \frac{\partial L}{\partial X(j_1,j_2,j_3)} = \langle \frac{\partial L}{\partial Y}, W(:,:,:,j_1,j_2,j_3)\rangle$

$\therefore \frac{\partial L}{\partial X} = W^T \frac{\partial L}{\partial Y}$

43.5 $Y=f(X)=WX$,

$W\in R^{(m_1\times m_2\times m_3)\times(n_1\times n_2\times n_3)}$

$X\in R^{(n_1\times n_2\times n_3)\times b}$, $Y\in R^{(m_1\times m_2\times m_3)\times b}$

given $\frac{\partial L}{\partial Y}$, here $L=L(Y)\in R$.

$\Rightarrow \frac{\partial L}{\partial W} = \frac{\partial L}{\partial Y}\cdot X^T$ (see 27)

$\frac{\partial L}{\partial X} = W^T \frac{\partial L}{\partial Y}$ (see 33)

44. 2D locally connected function:

$Y = f(X) = K \odot X$

$K \in R^{(m_1 \times m_2) \times (n_{k_1} \times n_{k_2} \times n_3)}$

$X \in R^{(n_1 \times n_2 \times n_3)}$, $Y \in R^{m_1 \times m_2}$

$\Rightarrow Y(i_1, i_2) = \langle K(i_1, i_2, :, :, :), X' \rangle$

$X' = X(i_1 : i_1 + n_{k_1}, i_2 : i_2 + n_{k_2}, \cdot)$

(~~这里的参数名字来自~~ cnn)

44.3 if $K \in R^{(m_1 \times m_2 \times m_3) \times (n_{k_1} \times n_{k_2} \times n_3)}$

$\Rightarrow Y \in R^{m_1 \times m_2 \times m_3}$

$Y(i_1, i_2, i_3) = \langle K(i_1, i_2, i_3, :, :, :), X' \rangle$

$X' = X(i_1 : i_1 + n_{k_1}, i_2 : i_2 + n_{k_2}, \cdot)$

44.4

矩阵乘 ⊃ 卷积 ⊃ 求导.

44.5 已知44.3, given $\frac{\partial L}{\partial Y}$, $L = L(Y) \in R$,

$\Rightarrow \frac{\partial L}{\partial K}, \frac{\partial L}{\partial X}$.

solve:

① $\because \frac{\partial Y(i_1, i_2, i_3)}{\partial K(i_1, i_2, i_3, j_1, j_2, j_3)}$

$= X'(j_1, j_2, j_3)$ (见44.3的定义)

$= X((i_1 - 1) + j_1, (i_2 - 1) + j_2, j_3)$

$\therefore \frac{\partial L}{\partial K(i_1, i_2, i_3, j_1, j_2, j_3)}$

$= \frac{\partial L}{\partial Y(i_1, i_2, i_3)} X((i_1 - 1) + j_1, (i_2 - 1) + j_2, j_3)$

② ~~[illegible]~~

② $\frac{\partial L}{\partial X(j_1, j_2, j_3)}$

$$= \sum_{\substack{i_1, i_2, i_3 \\ j_{k_1}, j_{k_2} \\ s.t.\ (i_1 - 1) + j_{k_1} = j_1 \\ (i_2 - 1) + j_{k_2} = j_2}} \frac{\partial L}{\partial Y(i_1, i_2, i_3)} K(i_1, i_2, i_3, j_{k_1}, j_{k_2}, j_3)$$

44.6 (解法2 of ~~[illegible]~~ 44.3)

~~solve~~ $Y = f(X) = K \odot X$,

$K \in R^{(m_1 \times m_2 \times m_3) \times (n_{k_1} \times n_{k_2} \times n_3)}$

$X \in R^{n_1 \times n_2 \times n_3}$, $Y \in R^{m_1 \times m_2 \times m_3}$

$\Rightarrow Y = ?$

solve:

① define $X' \in R^{(n_{k_1} \times n_{k_2} \times n_3) \times (m_1 \times m_2)}$

$X'(j_1, j_2, j_3, i_1, i_2)$

$= X(i_1 + j_1 - 1, i_2 + j_2 - 1, j_3)$

here $m_1 = n_1 - n_{k_1} + 1$, $m_2 = n_2 - n_{k_2} + 1$.

② ~~[illegible]~~

$\because Y(i_1, i_2, \cdot)$

$= K(i_1, i_2, \cdot, \cdot, \cdot, \cdot) \cdot X'(\cdot, \cdot, \cdot, i_1, i_2)$

$\forall i_1 \in \{1, \cdots, m_1\}$, $i_2 \in \{1, \cdots, m_2\}$

$\therefore$ Y便得到.

44.7. (解法2 of 44.5).

已知 $Y = K \odot X$,

$K \in R^{(m_1 \times m_2 \times m_3) \times (n_{k_1} \times n_{k_2} \times n_3)}$

$X \in R^{n_1 \times n_2 \times n_3}$, $Y \in R^{m_1 \times m_2 \times m_3}$

given $\frac{\partial L}{\partial Y}$, $L = L(Y) \in R$,

$\Rightarrow \frac{\partial L}{\partial K}, \frac{\partial L}{\partial X}$.

solve:

① define $X' \in R^{(n_{k_1} \times n_{k_2} \times n_3) \times (m_1 \times m_2)}$

$X'(j_1, j_2, j_3, i_1, i_2) = X(i_1 + j_1 - 1, i_2 + j_2 - 1, j_3)$

$\therefore Y(i_1, i_2, \cdot) = K(i_1, i_2, \cdot, :, :, :) \cdot X'(:, :, :, i_1, i_2)$

② $\therefore \frac{\partial L}{\partial K(i_1, i_2, :, :, :)} = \frac{\partial L}{\partial Y(i_1, i_2, :)} \cdot X'^T(:, :, :, i_1, i_2)$

$\forall i_1, i_2$. (一维列向量乘以行向量)

③ $\therefore \frac{\partial L}{\partial X'(:, :, :, i_1, i_2)} = K^T(i_1, i_2, :, :, :, :) \frac{\partial L}{\partial Y(i_1, i_2, :)}$

$\forall i_1 \in \{1, \cdots, m_1\}$, $i_2 \in \{1, \cdots, m_2\}$

$\therefore$ we get $\frac{\partial L}{\partial X'}$

④ 令 $\frac{\partial L}{\partial X} = 0 \in R^{n_1 \times n_2 \times n_3}$

$\forall j_1 \in \{1, \cdots, n_{k_1}\}$, $j_2 \in \{1, \cdots, n_{k_2}\}$, $j_3 \in \{1, \cdots, n_3\}$

$i_1 \in \{1, \cdots, m_1\}$, $i_2 \in \{1, \cdots, m_2\}$

$\frac{\partial L}{\partial X(i_1 + j_1 - 1, i_2 + j_2 - 1, j_3)} \mathrel{+}= \frac{\partial L}{\partial X'(j_1, j_2, j_3, i_1, i_2)}$

$\therefore$ we get $\frac{\partial L}{\partial X}$.

45. 2D tiled convolution.

$Y = f(X) = K \odot X$.

$K \in R^{(m_{t_1} \times m_{t_2}) \times (n_{k_1} \times n_{k_2} \times n_3)}$

$X \in R^{n_1 \times n_2 \times n_3}$, $Y \in R^{m_1 \times m_2}$

$\Rightarrow Y(i_1, i_2) = \langle K(i_1 \% m_{t_1} + 1, i_2 \% m_{t_2} + 1), X' \rangle$

here $X' = X(i_1 : i_1 + n_{k_1}, i_2 : i_2 + n_{k_2}, \cdot)$

45.3 if $K \in R^{(m_{t_1} \times m_{t_2} \times m_3) \times (n_{k_1} \times n_{k_2} \times n_3)}$

$\Rightarrow Y \in R^{m_1 \times m_2 \times m_3}$

$Y(i_1, i_2, i_3) = \langle K(i_1 \% m_{t_1} + 1, i_2 \% m_{t_2} + 1, i_3, :, :, :), X' \rangle$

here $X' = X(i_1 : i_1 + n_{k_1}, i_2 : i_2 + n_{k_2}, \cdot)$

46. 2D strided convolution:

$Y = f(X) = K \odot X$

$K \in R^{n_{k_1} \times n_{k_2} \times n_3}$,

$X \in R^{n_1 \times n_2 \times n_3}$, $Y \in R^{m_1 \times m_2}$

$\Rightarrow Y(i_1, i_2) = \langle K, X' \rangle$

here $X' = X(i_1' : i_1' + n_{k_1}, i_2' : i_2' + n_{k_2}, \cdot)$

$i_1' = (i_1 - 1) n_{s_1} + 1$, $i_2' = (i_2 - 1) n_{s_2} + 1$

n_{s_1}, n_{s_2} 是 stride.

46.3 if $K \in R^{(m_3) \times (n_{k_1} \times n_{k_2} \times n_3)}$

$\Rightarrow Y \in R^{m_1 \times m_2 \times m_3}$

$Y(i_1, i_2, i_3) = \langle K(i_3, :, :, :), X' \rangle$, X' 同上.

46.4 包含关系 of 各种函数.

average pooling $\subset$ convolution $\subset$ tiled convolution $\subset$ locally connected $\subset$ fully connected.

46.5. $Y = f(X) = K \odot X$,

$K \in \mathbb{R}^{m_3 \times (n_{k_1} \times n_{k_2} \times n_3)}$

$X \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $Y \in \mathbb{R}^{m_1 \times m_2 \times m_3}$

given $\frac{\partial L}{\partial Y}$, $\Rightarrow \frac{\partial L}{\partial K}, \frac{\partial L}{\partial X}$

solve: ①:

$\because Y(i_1,i_2,i_3) = \langle K(i_3,:,:,:), X' \rangle$

$X' = X(i_1' : i_1' + n_{k_1}, \; i_2' : i_2' + n_{k_2}, \; :)$

$i_1' = (i_1-1)n_{s_1}+1, \quad i_2' = (i_2-1)n_{s_2}+1$.

$$\therefore \frac{\partial Y(i_1,i_2,i_3)}{\partial K(i_3,j_1,j_2,j_3)} = X'(j_1,j_2,j_3) = X((i_1-1)n_{s_1}+j_1, (i_2-1)n_{s_2}+j_2, j_3)$$

$$\therefore \frac{\partial L}{\partial K(i_3,j_1,j_2,j_3)} = \sum_{i_1,i_2} \frac{\partial L}{\partial Y(i_1,i_2,i_3)} \frac{\partial Y(i_1,i_2,i_3)}{\partial K(i_3,j_1,j_2,j_3)}$$

$$= \sum_{i_1,i_2} \frac{\partial L}{\partial Y(i_1,i_2,i_3)} X((i_1-1)n_{s_1}+j_1, (i_2-1)n_{s_2}+j_2, j_3)$$

$$② \quad \frac{\partial L}{\partial X(j_1,j_2,j_3)} = \sum_{\substack{i_1,i_2,i_3 \\ j_{k_1}, j_{k_2} \\ \text{s.t. } (i_1-1)n_{s_1}+j_{k_1}=j_1 \\ (i_2-1)n_{s_2}+j_{k_2}=j_2}} \frac{\partial L}{\partial Y(i_1,i_2,i_3)} K(i_3, j_{k_1}, j_{k_2}, j_3)$$

46.6. 解法2 of 46.5

solve: $\because$ ~~strided~~ convolution $\subset$ locally connected

$\therefore$ 可参考44.7 给出的求解过程.

① define $X' \in \mathbb{R}^{(n_{k_1} \times n_{k_2} \times n_3) \times (m_1 \times m_2)}$

$X'(j_1,j_2,j_3,i_1,i_2)$

$= X((i_1-1)n_{s_1}+j_1, (i_2-1)n_{s_2}+j_2, j_3)$

$\therefore Y(i_1,i_2,:) = K \cdot X'(:,:,:,i_1,i_2)$

若 $Y'(i_3,i_1,i_2) = Y(i_1,i_2,i_3)$, //这样定义$Y'$.

则 $Y' = KX'$, 这里我们将Y转成Y'

将 $\frac{dL}{dY}$ 转成 $\frac{dL}{dY'}$,

但为了方便, 仍用记号Y, 而非Y'.

(以后对Y定义时, 用Y'的定义).

② $\because Y = KX'$, (here $\overset{m_1 \times m_2}{\boxed{m_3}} = \square \times \boxed{\quad}$)

$\therefore \frac{\partial L}{\partial K} = \frac{\partial L}{\partial Y} \cdot X'$

$\frac{\partial L}{\partial X'} = K^T \frac{\partial L}{\partial Y}$

③ 令 $\frac{\partial L}{\partial X} = 0 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$

$\forall j_1, j_2, j_3, i_1, i_2$

$$\frac{\partial L}{\partial X((i_1-1)n_{s_1}+j_1, (i_2-1)n_{s_2}+j_2, j_3)} \mathrel{+}= \frac{\partial L}{\partial X'(j_1,j_2,j_3,i_1,i_2)}$$

$\therefore$ we get $\frac{\partial L}{\partial X}$.

47. 2D convolution:

$Y = f(X) = K \odot X$,

$K \in R^{n_{k_1} \times n_{k_2} \times n_3}$, $X \in R^{n_1 \times n_2 \times n_3}$,

$Y \in R^{m_1 \times m_2}$

$\Rightarrow Y(i_1, i_2) = \langle K, X' \rangle$

$X' = X(i_1 : i_1 + n_{k_1}, i_2 : i_2 + n_{k_2}, \cdot)$

47.3 if $K \in R^{(m_3) \times (n_{k_1} \times n_{k_2} \times n_3)}$

$\Rightarrow Y \in R^{m_1 \times m_2 \times m_3}$

$Y(i_1, i_2, i_3) = \langle K(i_3, :, :, :), X' \rangle$

$X' = X(i_1 : i_1 + n_{k_1}, i_2 : i_2 + n_{k_2}, \cdot)$

48. average pooling function:

$Y = f(X) = K \odot X$,

$K = \frac{1}{n_{k_1} \cdot n_{k_2}} \in R$, $X \in R^{n_1 \times n_2 \times n_3}$

$Y \in R^{m_1 \times m_2 \times m_3}$, with $m_3 = n_3$

$\Rightarrow Y(i_1, i_2, i_3) = \frac{1}{n_{k_1} \cdot n_{k_2}} \operatorname{sum} X'$

$i_1' = (i_1 - 1) n_{s_1} + 1$, $i_2' = (i_2 - 1) n_{s_2} + 1$

$X' = X(i_1' : i_1' + n_{k_1}, i_2' : i_2' + n_{k_2}, i_3)$,

sumX'是指所有元素求和.

48.3: 已知48中的条件, 且 given $\frac{\partial L}{\partial Y}$,

$\Rightarrow \frac{\partial L}{\partial X}$.

solve: $\frac{\partial L}{\partial X(j_1, j_2, j_3)} = \sum_{\substack{i_1, i_2, \\ j_{k_1}, j_{k_2} \\ \text{s.t. } (i_1-1) n_{s_1} + j_{k_1} = j_1 \\ (i_2 - 1) n_{s_2} + j_{k_2} = j_2}} \frac{\partial L}{\partial Y(i_1, i_2, i_3)} \cdot \frac{1}{n_{k_1} \cdot n_{k_2}}$, $j_3 = i_3$.

49. max pooling function:

$Y = f(X)$,

$X \in R^{n_1 \times n_2 \times n_3}$, $Y \in R^{m_1 \times m_2 \times m_3}$, $n_3 = m$

$\Rightarrow Y(i_1, i_2, i_3) = \max X'$

$X' = X(i_1' : i_1' + n_{k_1}, i_2' : i_2' + n_{k_2}, i_3)$

$i_1' = (i_1 - 1) n_{s_1} + 1$, $i_2' = (i_2 - 1) n_{s_2} + 1$

49.3 已知49, 且 given $\frac{\partial L}{\partial Y}$, $L = L(Y) \in R$,

$\Rightarrow \frac{\partial L}{\partial X}$.

solve:

$\frac{\partial L}{\partial X(j_1, j_2, j_3)}$

$= \sum_{\substack{i_1, i_2 \\ j_{k_1}, j_{k_2} \\ \text{s.t. } (i_1 - 1) n_{s_1} + j_{k_1} = j_1 \\ (i_2 - 1) n_{s_2} + j_{k_2} = j_2}} \frac{\partial L}{\partial Y(i_1, i_2, i_3)} \cdot 1_{(j_1, j_2, j_3) = q(i_1, i_2, i_3)}$, $i_3 = j_3$.

here $q(i_1, i_2, i_3) = \operatorname{argmax}_{(j_1, j_2, j_3) \in S} X(j_1, j_2, j_3)$

$S = \{(i_1 - 1) n_{s_1} + j_{k_1}, (i_2 - 1) n_{s_2} + j_{k_2}, i_3 \mid \forall j_{k_1}, j_{k_2}$

50. logistic function

$$\sigma(x)=\frac{1}{1+e^{-x}}$$

50.3 $\log\frac{\sigma(x)}{1-\sigma(x)}=x$

50.4 $\sigma'(x)=\sigma(x)(1-\sigma(x))$

51. $x \xrightarrow{\sigma} y \xrightarrow{L} L$

here $L(y)=-t\log y-(1-t)\log(1-y)$

$t\in\{0,1\}$

求: $\frac{\partial L}{\partial x}$

solve: $\frac{\partial L}{\partial x}=\frac{\partial y}{\partial x}\frac{\partial L}{\partial y}=y(1-y)\cdot\frac{y-t}{y(1-y)}$

$=y-t$

52. softmax function:

$$y=f(x)=\frac{1}{\sum_{i=1}^{n}e^{x_i}}e^{x}=\frac{e^{x}}{\Sigma}$$

$x\in R^n$, $x=(x_1,\cdots,x_n)$

$e^{x}=(e^{x_1},\cdots,e^{x_n})$, $\Sigma:=\sum_{i=1}^{n}e^{x_i}$

52.3 求 f'

solve: if $i=j$:

$$\frac{\partial y_i}{\partial x_j}=\frac{\partial(e^{x_i}/\Sigma)}{\partial x_j}=y_i(1-y_i)$$

if $i\neq j$:

$$\frac{\partial y_i}{\partial x_j}=\frac{\partial(e^{x_i}/\Sigma)}{\partial x_j}=-y_iy_j$$

53.

函数关系:

$$x \xrightarrow{softmax} y \xrightarrow{L} L$$

here $L=-\sum_{i=1}^{n}t_i\log y_i$,

$\sum_i t_i=1$, $t_i\in\{0,1\}$, $x,y\in R^n$

求: $\frac{\partial L}{\partial x}$.

solve:

$$\begin{aligned}\frac{\partial L}{\partial x_j}&=-\sum_{i=1}^{n}t_i\frac{\partial\log y_i}{\partial x_j}\\&=-\sum_i t_i\frac{1}{y_i}\frac{\partial y_i}{\partial x_j}\\&=-\sum_{i,i\neq j}t_i\frac{1}{y_i}\frac{\partial y_i}{\partial x_j}-t_j\frac{1}{y_j}\frac{\partial y_j}{\partial x_j}\\&=\sum_{i,i\neq j}t_iy_j-t_j(1-y_j)\\&=y_j\sum_i t_i-t_j\\&=y_j-t_j\end{aligned}$$

54. $y, t \in R^n$, $L = \|y-t\|^2$

$\Rightarrow \frac{\partial L}{\partial y}$

solve: $\because L = (y-t)^T(y-t)$

$\therefore \frac{\partial L}{\partial y} = 2(y-t)$

55. $y, t \in R^n$,

$L = I(y,t)\|y-t\|^2$

$$I(y,t) = \begin{cases} 1, & \|y\| \geq \|t\| \\ 0, & \|y\| < \|t\| \end{cases}$$

$\Rightarrow \frac{\partial L}{\partial y}$

solve: $$\frac{\partial L}{\partial y} = \begin{cases} 2(y-t), & \|y\| \geq \|t\| \\ 0, & \|y\| < \|t\| \end{cases}$$

这里 $\frac{\partial L}{\partial y}$ 在 $\|y\| = \|t\|$ 时可导的严谨性暂忽略.

56. $y, t_1, t_2 \in R^n$

$L = I(y,t)\|y-t\|^2$

$$I(y,t) = \begin{cases} 1, & \|y\| \geq \|t\| \\ 0, & \|y\| < \|t\| \end{cases}$$

$$t = \begin{cases} t_1, & \text{if } \|y-t_1\| \geq \|y-t_2\| \\ t_2, & \text{otherwise} \end{cases}$$

$\Rightarrow \frac{\partial L}{\partial y}$

solve: ① if $\|y-t_1\| \geq \|y-t_2\|$:

$$\frac{\partial L}{\partial y} = \begin{cases} 2(y-t_1), & \text{if } \|y\| \geq \|t_1\| \\ 0, & \text{if } \|y\| < \|t_1\| \end{cases}$$

② if $\|y-t_1\| < \|y-t_2\|$:

$$\frac{\partial L}{\partial y} = \begin{cases} 2(y-t_2), & \text{if } \|y\| \geq \|t_2\| \\ 0, & \text{if } \|y\| < \|t_2\| \end{cases}$$

57. $x \xrightarrow{\sigma} y \xrightarrow{L} L$, $k_1, k_2 \in R^+$

$L(y) = I(k_1, k_2)(-t\log y - (1-t)\log(1-y))$

$y = \sigma(x) = \frac{1}{1+e^{-x}}$

$$I(k_1,k_2) = \begin{cases} 1, & k_1 \geq 2k_2 \text{ or } k_1 \leq \frac{1}{2}k_2 \\ 0, & \frac{1}{2}k_2 < k_1 < 2k_2 \end{cases}$$

$$t = \begin{cases} 1, & \text{if } k_1 \geq 2k_2 \\ 0, & k_1 \leq \frac{1}{2}k_2 \\ 2, & \frac{1}{2}k_2 < k_1 < 2k_2 \end{cases}$$

$\Rightarrow \frac{\partial L}{\partial x}$

solve: $\because \frac{\partial L}{\partial x} = I(k_1,k_2)(y-t)$ (see 51)

$$\therefore \frac{\partial L}{\partial x} = \begin{cases} y-1, & k_1 \geq 2k_2 \\ y, & k_1 \leq \frac{1}{2}k_2 \\ 0, & \frac{1}{2}k_2 < k_1 < 2k_2 \end{cases}$$

58. ① $x \xrightarrow{\sigma} y \xrightarrow{L} L$, $L(y) = -t\log y - (1-t)\log(1-y)$

$t \in [0,1]$, define $L = f_1(x)$

② $x \to y \xrightarrow{L} L$, $L(y) = \frac{1}{2}(y-t)^2$, $y = x$.

$t \in [0,1]$, define $L = f_2(x)$

$\Rightarrow f_1'(x) = f_2'(x) = y - t$

59. $L = I_1(y,t)(2 - t_w t_h) L_1 + I_2(y,t) L_2 + I_1(y,t) L_3$

$$I_2(y,t) = \begin{cases} 1, & \|y\| \geq 2\|t\|, \text{ or } \|y\| \leq \frac{1}{2}\|t\| \\ 0, & \text{otherwise} \end{cases}$$

$y = (y_x, y_y, y_w, y_h, y_c, y_{p_1}, \cdots, y_{p_n}) \in R^{5+n}$

define $S = \{x, y, w, h, c, p_1, \cdots, p_n\}$

$t = (\cdots, t_i, \cdots) \in R^{5+n}$, $i \in S$

$x \in (\cdots, x_i, \cdots) \in R^{5+n}$, $i \in S$.

$$t_c = \begin{cases} 1, & \text{if } \|y\| \geq 2\|t\| \\ 0, & \|y\| \leq \frac{1}{2}\|t\| \\ 2, & \frac{1}{2}\|t\| < \|y\| < 2\|t\| \end{cases}$$

$$\begin{cases} y_i = \sigma(x_i), & i \in S,\ i \neq w,\ i \neq h \\ y_i = x_i, & i = w \text{ or } i = h \end{cases}$$

$\Rightarrow \frac{\partial L}{\partial x}$

define $H_b(p,q) = -p\log q - (1-p)\log(1-q)$

$p, q \in R$

$L_1 = L_{11} + L_{12}$,

$L_{11} = H_b(t_x, y_x) + H_b(t_y, y_y)$

$L_{12} = \frac{1}{2}\left((y_h - t_h)^2 + (y_w - t_w)^2\right)$

$L_2 = H_b(t_c, y_c)$

$$L_3 = \begin{cases} L_{31}, & \text{if draw } v \text{ from } N(0,1), \text{ and } v < 0 \\ L_{32}, & \text{if draw } v \text{ from } N(0,1), \text{ and } v \geq 0 \end{cases}$$

$L_{31} = \sum_{i=1}^{n} H_b(t_{p_i}, y_{p_i})$, $L_{32} = H_b(t_{p_r}, y_{p_r})$

$$t_{p_i} = \begin{cases} 1, & i = r, \text{ here } r \in N, \text{ is constant}, r \leq n. \\ 0, & \text{otherwise} \end{cases}$$

$$I_1(y,t) = \begin{cases} 1, & \|y\| \geq 2\|t\| \\ 0, & \|y\| < 2\|t\| \end{cases}$$

solve:

$\because \frac{\partial L_1}{\partial x_i} = y_i - t_i$, $i \in \{x, y, w, h\}$

$\frac{\partial L_2}{\partial x_c} = y_c - t_c$.

$\frac{\partial L_{31}}{\partial x_{p_i}} = y_{p_i} - t_{p_i}$, $i = 1, \cdots, n$

$$\frac{\partial L_{32}}{\partial x_{p_i}} = \begin{cases} 0, & i \neq r \\ y_{p_i} - t_{p_i}, & i = r. \end{cases}$$

$\therefore$ ① $$\frac{\partial L}{\partial x_i} = \begin{cases} (2 - t_w t_h)(y_i - t_i), & \text{if } \|y\| \geq 2\|t\| \\ 0, & \text{otherwise}, \end{cases}$$

$i \in \{x, y, w, h\}$

② $$\frac{\partial L}{\partial x_c} = \begin{cases} y_c - 1, & \|y\| \geq 2\|t\| \\ y_c, & \|y\| \leq \frac{1}{2}\|t\| \\ 0, & \text{otherwise} \end{cases}$$

③ draw v from $N(0,1)$,

if $v < 0$; $\forall i$, $$\frac{\partial L}{\partial x_{p_i}} = \begin{cases} y_{p_i} - t_{p_i}, & \|y\| \geq 2\|t\| \\ 0, & \|y\| < 2\|t\| \end{cases}$$

if $v \geq 0$, $$\frac{\partial L}{\partial x_{p_i}} = \begin{cases} y_{p_i} - t_{p_i}, & \|y\| \geq 2\|t\| \\ 0, & \text{otherwise} \end{cases}$$

$i = r$. $\frac{\partial L}{\partial x_{p_i}} = 0$, for

60. $y = Wx + b$

$\Rightarrow \frac{\partial y}{\partial x} = W$

61. $y = UVWx$

$\Rightarrow \frac{\partial y}{\partial x} = UVW$

62. $y = UVWx, \quad x \in R^n, y \in R$

$\Rightarrow \frac{\partial y_i}{\partial x_j}$

62.3 method 1.

solve:

$\because \frac{\partial y}{\partial x} = UVW = A$

$\therefore \frac{\partial y_i}{\partial x_j} = A_{ij}$

let u^T be i-th row of U,

w be j-th column of W.

$\therefore \frac{\partial y_i}{\partial x_j} = A_{ij} = u^T V w$

62.4 method 2.

solve:

$\because y \xleftarrow{U} a \xleftarrow{V} b \xleftarrow{W} x$

$\therefore \frac{\partial y_i}{\partial x_j} = \sum_{l_1} \frac{\partial y_i}{\partial b_{l_1}} \frac{\partial b_{l_1}}{\partial x_j}$

$= \sum_{l_1} \left(\sum_{l_2} \frac{\partial y_i}{\partial a_{l_2}} \frac{\partial a_{l_2}}{\partial b_{l_1}} \right) \frac{\partial b_{l_1}}{\partial x_j}$

$= \sum_{l_1} \sum_{l_2} \frac{\partial y_i}{\partial a_{l_2}} \frac{\partial a_{l_2}}{\partial b_{l_1}} \frac{\partial b_{l_1}}{\partial x_j}$

$= \sum_{l_1} \sum_{l_2} \cancel{A_{ij}} \, U_{i,l_2} V_{l_2,l_1} W_{l_1,j}$

(water flow)

63. $y = f(Wx+b)$

f is element-wise, i.e.,

$f(x) = (f(x_1), f(x_2), \cdots, f(x_n))$

$\forall x \in R^n$

$\Rightarrow \frac{\partial y}{\partial x} = \frac{\partial y}{\partial (Wx)} \cdot \frac{\partial (Wx)}{\partial x} = F' W$

64. $y^{(i)} = f_i(W^{(i)} y^{(i-1)} + b^{(i)})$

$i = 1, 2, 3.$

$y = y^{(3)}, \quad x = y^{(0)},$

f_i is element-wise.

$\Rightarrow$ ① $\frac{\partial y}{\partial x}$, ② $\frac{\partial y_i}{\partial x_j}$

solve:

① $\frac{\partial y}{\partial x} = \frac{\partial y^{(3)}}{\partial y^{(2)}} \frac{\partial y^{(2)}}{\partial y^{(1)}} \frac{\partial y^{(1)}}{\partial y^{(0)}}$

$= F_3' W^{(3)} F_2' W^{(2)} F_1' W^{(1)}$

$= A.$

where $F' = \begin{pmatrix} f' & & \\ & f' & \\ & & \ddots \end{pmatrix}$

② define u^T i-th row of $W^{(3)}$

v j-th column of $W^{(1)}$

$\therefore \frac{\partial y_i}{\partial x_j} = A_{ij} = f_3' u^T F_2' W^{(2)} F_1' v$

65. $y^{(i)} = f_i(W^{(i)} y^{(i-1)} + b^{(i)})$

$i = 1, 2, 3$

$y = y^{(3)}, \quad x = y^{(0)},$

f_i is element-wise.

$J = J(y) \in R$

$x \to y^{(1)} \to y^{(2)} \to y \to J$

$v_{y^{(i)}} := \frac{\partial J}{\partial y^{(i)}}$

$\Rightarrow$ ① $\frac{\partial v_x}{\partial v_y}$ ② $\frac{\partial v_{x_j}}{\partial v_{y_i}}$

solve:

① $\because y^{(i)} = f_i(W^{(i)} y^{(i-1)} + b^{(i)})$

$\therefore v_{y^{(i-1)}} = v_{y^{(i)}} F_i' W^{(i)}$

$\therefore v_x \leftarrow v_{y^{(1)}} \leftarrow v_{y^{(2)}} \leftarrow v_y$

$\therefore v_x = v_y F_3' W^{(3)} F_2' W^{(2)} F_1' W^{(1)} = v_y A$

$\therefore \frac{\partial v_x}{\partial v_y} = A^T$

③ $\dfrac{\partial v_{x_j}}{\partial v_{y_i}} = (A^T)_{ji} = A_{ij}$

define u^T i-th row of $W^{(3)}$

v j-th column of $W^{(1)}$

$\therefore \dfrac{\partial v_{x_j}}{\partial v_{y_i}} = A_{ij} = f_3' u^T F_2' W^{(2)} F_1' v$

66 $y^{(i)} = f_i(W^{(i)} y^{(i-1)} + b^{(i)})$

$i = 1, 2, \cdots, l$

$y = y^{(l)}$, $x = y^{(0)}$, f_i element-wise

$J = J(y) \in \mathbb{R}$

$v_{y^{(i)}} := \dfrac{\partial J}{\partial y^{(i)}}$

$\Rightarrow$ ① $\dfrac{\partial y}{\partial x} = A$

② $\dfrac{\partial v_x}{\partial v_y} = A^T$

③ $\dfrac{\partial y_i}{\partial x_j} = \dfrac{\partial v_{x_j}}{\partial v_{y_i}} = A_{ij}$

where $A = \prod_{i=l}^{i=1} F_i' W^{(i)}$

67. $y^{(i)} = f_i(W^{(i)} y^{(i-1)} + b^{(i)})$

$i = 1, 2, \cdots, l$, $l > 1$

$y = y^{(l)}$, $x = y^{(0)}$, f_i element-wise

$J = J(y) \in \mathbb{R}$

$v_{net^{(i)}} := \dfrac{\partial J}{\partial net^{(i)}}$

$net^{(i)} := W^{(i)} y^{(i-1)} + b^{(i)}$

$\Rightarrow$ ① $\dfrac{\partial net^{(l)}}{\partial net^{(1)}} = B$

② $\dfrac{\partial v_{net^{(1)}}}{\partial v_{net^{(l)}}} = B^T$

③ $\dfrac{\partial net_i^{(l)}}{\partial net_j^{(1)}} = \dfrac{\partial v_{net_j^{(1)}}}{\partial v_{net_i^{(l)}}} = B_{ij}$

where $B = \prod_{i=l}^{i=2} W^{(i)} F_{i-1}'$

Proof: let $F_i' = I$, $W^{(1)} = I$

$\therefore \dfrac{\partial net^{(l)}}{\partial net^{(1)}} = \dfrac{\partial y}{\partial x} = A = B$ (using 66)

Similarly, we can prove others.