1、(1) $E_{y|x} = w_0 + W^T X$

误差 $S = \sum_i (E_{y|x} - \overset{gt}{y_i})^2 = \sum_i (w_0 + W^T x - y_i)^2$

$\frac{ds}{dW} = 2\sum_i^n (w_0 + W^T X - y_i) X_i = 0$

$\Rightarrow (w_0 + WX_1 - y_1) + (w_0 + WX_2 - y_2) + \cdots + (w_0 + WX_n - y_n) = 0$

$\Rightarrow n w_0 + W(X_1 + \cdots + X_n) - (y_1 + \cdots + y_n) = 0$

$\Rightarrow n w_0 + W \sum_i^n X_i^T - \sum_i^n y_i = 0$

$w_0 = \frac{1}{n}\sum_i y_i - \frac{1}{n}\sum_i X_i^T W = \bar{y} - \bar{X}^T W$

(2)

loss function can transfor to

$\leftarrow$ $\ell(h, x, y) = (h(x) - y)^2$, w.r.t

$h(x_i) = Wx_i + b$

$(w^*, b^*) = \underset{(w,b)}{\arg\min} \sum_{i=1}^m (h(x_i) - y_i)^2$

$= \underset{(w,b)}{\arg\min} \sum_{i=1}^m (y_i - WX_i - b)^2$

求解 w, b 使 $E_{(w,b)} = \sum_{i=1}^m (y_i - WX_i - b)^2$ 最小化

$\frac{\partial E(w,b)}{\partial W} = \cancel{2(\sum_{i=1}^m (y_i - wx_i - b) \cdot x_i} = 2(W\sum_{i=1}^m X_i^2 - \sum_{i=1}^m (y_i - b) X_i)$ ①

$\frac{\partial E(w,b)}{\partial b} = 2(mb - \sum_{i=1}^m (y_i - wX_i))$ ②

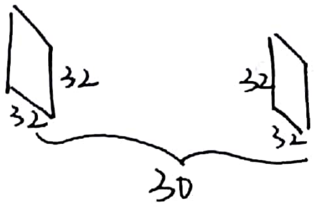
令①、②式得 0

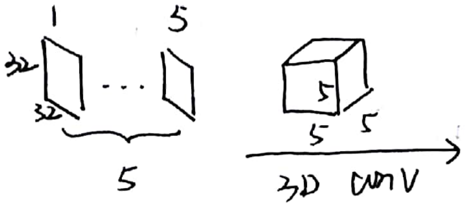
$$W = \frac{\sum_{i=1}^m y_i (X_i - \bar{x})}{\sum_{i=1}^m X_i^2 - \frac{1}{m}(\sum_{i=1}^m X_i)^2}$$

$$b = \frac{1}{m}\sum_{i=1}^m (y_i - wX_i)$$

2. (1)

input video





$\frac{32-5}{2}+1 = 14$.

S=2

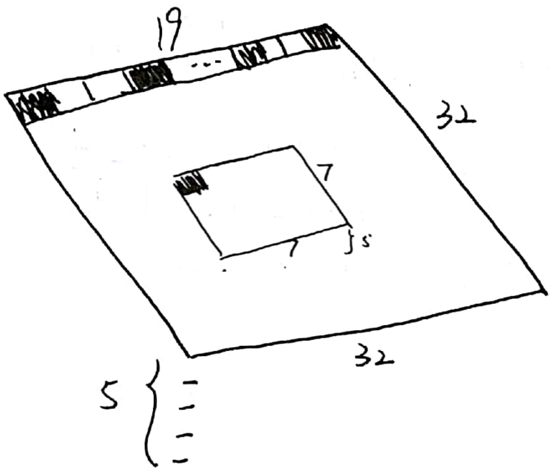
temporal stride = 4

frame: 1 ~ 5
5 ~ 9
9 ~ 13
13 ~ 17
17 ~ 21
21 ~ 25
25 ~ 29
29 ~ 30

$8 \times 14 \times 14 \times 64$ (no padding)

if we keep padding = 5 $\quad \frac{32-5+2\times5}{2}+1 = 19$ output size: $8\times19\times19\times64$

(2)



stride=1 $\quad rf = 19$

感受野: ~~rf=(rf−1)×s+k=(7−1)×1+7=13~~

$(W-k+2p)/s+1 = W$ 即尺寸不变 32×32

$\uparrow$ 19 $\quad \uparrow$ 9 $\quad \uparrow$ 1

$\Rightarrow p=9$

downsampling = upsampling = 2 (stride)

dilate rate=3

3.(1) $\mu_B = \frac{1}{m}\sum_{i=1}^{m} x_i$

$\sigma_B^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i-\mu_B)^2$

$\hat{x}_i = \frac{x_i-\mu_B}{\sqrt{\sigma_B^2+\epsilon}}$

$y_i = \gamma\hat{x}_i+\beta \equiv BN_{\gamma,\beta}(x_i)$

$\mu_B' = \frac{1}{m}\sum aWu = a\mu_B$

$\sigma_B' = \frac{1}{m}\sum_{i=1}^{m}(aWu-\mu_B')^2 = a^2\sigma_B^2$

$\hat{x}_i' = \frac{aWu-a\mu_B}{\sqrt{a^2\sigma_B^2+\epsilon}} = \hat{x}_i$

$\hat{y}_i' = BN_{\gamma,\beta}(aWu) = BN_{\gamma,\beta}(Wu)$

忽略 ϵ

$$BN(aWu) = \gamma\cdot\frac{aWu-\mu_B'}{\sqrt{a^2\sigma_B^2}} = \gamma\cdot\frac{Wu-\mu_B}{\sqrt{\sigma_B^2}} = BN(Wu)$$

$$\frac{\partial BN(aW)u}{\partial u} = \gamma\cdot\frac{aW}{\sqrt{\sigma_B'^2}} = \gamma\cdot\frac{aW}{\sqrt{a^2\sigma_B^2}} = \frac{\partial BN(Wu)}{\partial u}$$