1. 1) 令 $\hat{y}_i = w_0 + w^T x_i$ $\quad L(y_i, \hat{y}_i) = (\hat{y}_i - y_i)^2$

则 $E(L(y_i, \hat{y}_i)) = \frac{1}{n}\sum_{i=1}^{n}(w_0 + w^T x_i - y_i)^2$

令 $\frac{\partial E(L(y_i, \hat{y}_i))}{\partial w_0} = \frac{2}{n}(n w_0 + \sum_{i=1}^{n} x_i^T w - \sum_{i=1}^{n} y_i) = 0$

所以整理: $w_0 = \frac{1}{n}\sum_{i=1}^{n} y_i - \frac{1}{n}\sum_{i=1}^{n} x_i^T w = \bar{y} - \bar{x}^T w$

2) 构造新变量 $a_1, \dots a_n$.

$\min(\frac{1}{n}\sum_{i=1}^{n} a_i)$

$L(h, x, y) = \begin{cases} h(x) - y & h(x) \geq y \\ y - h(x) & h(x) < y \end{cases} \Rightarrow$ $\quad h(x_i) - y_i \leq a_i$

$h(x_i) - y_i \geq -a_i$

$i = 1, 2, \dots n$.

2. 1) 由题目已知 inputs = T × H × W × C1(RGB)

$T = 30 \quad H = 32 \quad W = 32$ ∵是3D卷积神经网络 ∴$C_1 = 3$ 三通道.

∴ inputs = 30 × 32 × 32 × 3

Kernels = $K_t \times K_h \times K_w \times C_1$(RGB) × C_2(数量)

由题目已知 $K_t = 5 \quad K_h = 5 \quad K_w = 5$ ∵是3D卷积神经网络 ∴C_1(RGB)=3 三通道

$C_2 = 64$

∴ Kernels = 5 × 5 × 5 × 3 × 64.

outputs = $t \times h \times w \times C_2$.

$t = (T - K_t + Pad_t)/S_t + 1$ S_t是时间的步长 由已知得 $S_t = 4$

$= (30 - 5 + Pad_t)/4 + 1 = (25 + Pad_t)/4 + 1$ 取 $Pad_t = 3$. 刚好能除尽4.

$= (25 + 3)/4 + 1 = 8$

$h = (H - K_h + Pad_h)/S_h + 1 = (32 - 5 + 1)/2 + 1 = 15$. $Pad_h = 1$ 刚好能除尽

$w = (W - K_w + Pad_w)/S_w + 1 = (32 - 5 + 1)/2 + 1 = 15$ $Pad_w = 1$ 刚好能除尽

∴ outputs = 8 × 15 × 15 × 64

2) 由公式 $\hat{K} = 1 + rate(K-1)$ 可得 k_t k_h k_w $rate = 3$

$\therefore k_t = 1+3(5-1) = 13$ $k_h = 1+3(7-1) = 19$ $k_w = 1+3(7-1) = 19$

再由第11的公式

即 $t = (T - k_t + Pad_t)/s_t + 1$ $h = (H - k_h + Pad_h)/s_h + 1$

$w = (W - k_w + Pad_w)/s_w + 1$

∵想要保持原来的分辨率 ∴ $t=30$ $h=32$ $w=32$ $c_1 = 3$ $c_2 = 1$

$t = (30 - 13 + Pad_t)/s_t + 1 = 30$

$h = (32 - 19 + Pad_h)/s_h + 1 = 32$

$w = (32 - 19 + Pad_h)/s_w + 1 = 32$

$\therefore outputs = t \times h \times w \times c_1 \times c_2 = 30 \times 32 \times 32 \times 3 \times 1$

downsampling = upsampling = rate = 3.

3. D

$$BN((\alpha W)u) = \frac{(\alpha W)u - E((\alpha W)u)}{\sqrt{Var((\alpha W)u)}}$$

$$= \frac{\alpha Wu - E(\alpha Wu)}{\sqrt{Var(\alpha Wu)}}$$

∵ α 是一个标量

∴ $E(\alpha Wu) = \alpha E(Wu)$

$Var(\alpha Wu) = \alpha^2 Var(Wu)$

∴ 上式 $= \frac{\alpha Wu - \alpha E(Wu)}{\sqrt{\alpha^2 Var(Wu)}}$

$= \frac{\alpha Wu - \alpha E(Wu)}{\alpha \sqrt{Var(Wu)}}$

$= \frac{Wu - E(Wu)}{\sqrt{Var(Wu)}} = BN(Wu)$

3) 由D可知

$BN((\alpha W)u) = BN(Wu)$

$\therefore \frac{\partial BN((\alpha W)u)}{\partial u} = \frac{\partial BN(Wu)}{\partial u}$