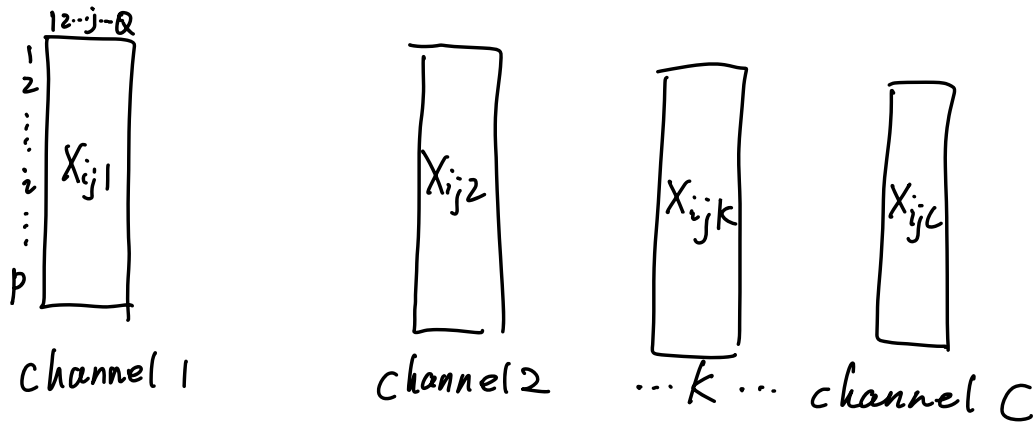


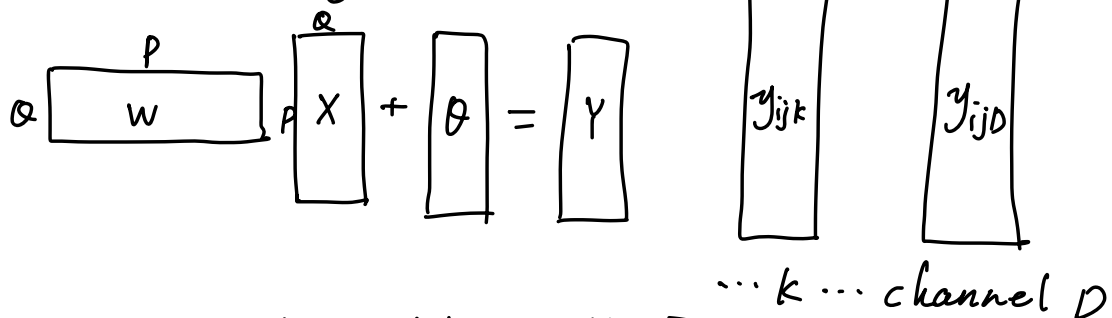
22-52-Q4

(a) Q: $X_{ijk} \rightarrow y_{ijk}$? parameters?

Solution ① understand



$X_{ijk} \rightarrow y_{ijk}$ full connect



$$W = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1i} & \dots & w_{1p} \\ w_{21} & w_{22} & \dots & w_{2i} & \dots & w_{2p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ w_{j1} & w_{j2} & \dots & w_{ji} & \dots & w_{jp} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ w_{Q1} & w_{Q2} & \dots & w_{Qi} & \dots & w_{Qp} \end{bmatrix} \begin{bmatrix} x_{11} & \dots & x_{1Q} \\ x_{21} & \dots & x_{2Q} \\ \vdots & \ddots & \vdots \\ x_{i1} & \dots & x_{iQ} \\ \vdots & \ddots & \vdots \\ x_{p1} & \dots & x_{pQ} \end{bmatrix} + b_k$$

$$y_{ijk} = \sum_{l=1}^C \sum_{j=1}^Q \sum_{i=1}^P w_{ji} x_{ijl} + b_k \quad (1 \leq k \leq D)$$

parameters $P \times Q \times C \times D + D$

(b) CNN ?



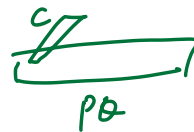
$$y_{i,j,k} = \sum_{u=-1}^1 \sum_{v=-1}^1 \sum_{l=1}^C w_{u,v,l,k} x_{i-u,j-v,l} + b_k$$

$$3 \times 3 \times C \times D + D = 9 \times C \times D + D$$

$$(c) y_{i,j,k} = \sum_{l=1}^C w_{l,k} x_{i,j,l} + b_k$$

$C \times D + D$ bias
 RLB D个特征图

$$(d) y_{i,k} = \sum_{l=1}^C w_{l,k} x_{i,l}$$



$C \times D$

$$(e) Y = XW$$



① Both architectures apply matrix multiplications to process input

② $Y = XW$ shows the CNN perform a linear transformation on the input features at each spatial position, identical across all position.

It is similar to the linear layers used in Transformer, where input are transformed via weight matrices