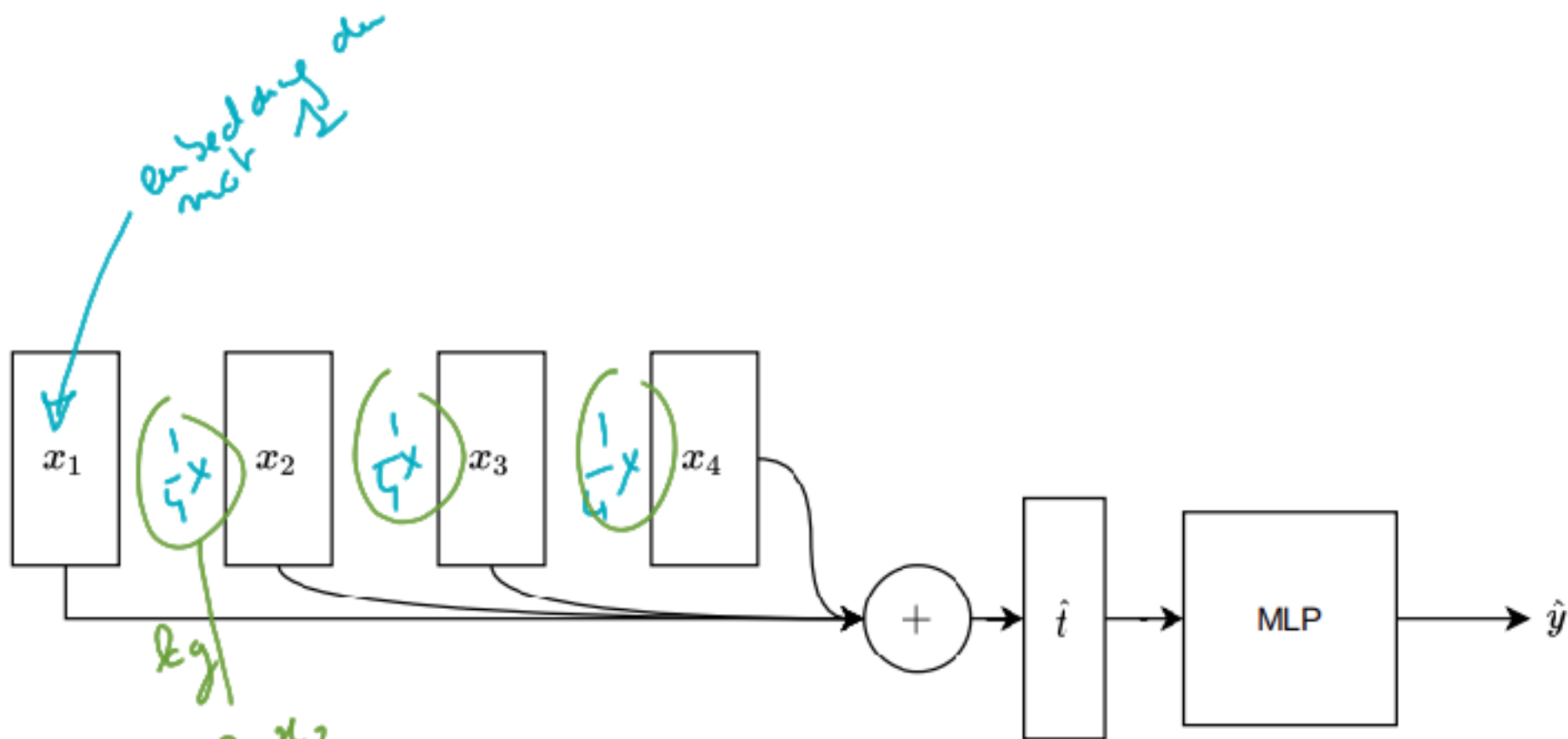


Q1



Q2

cte +  $q \cdot x_1$

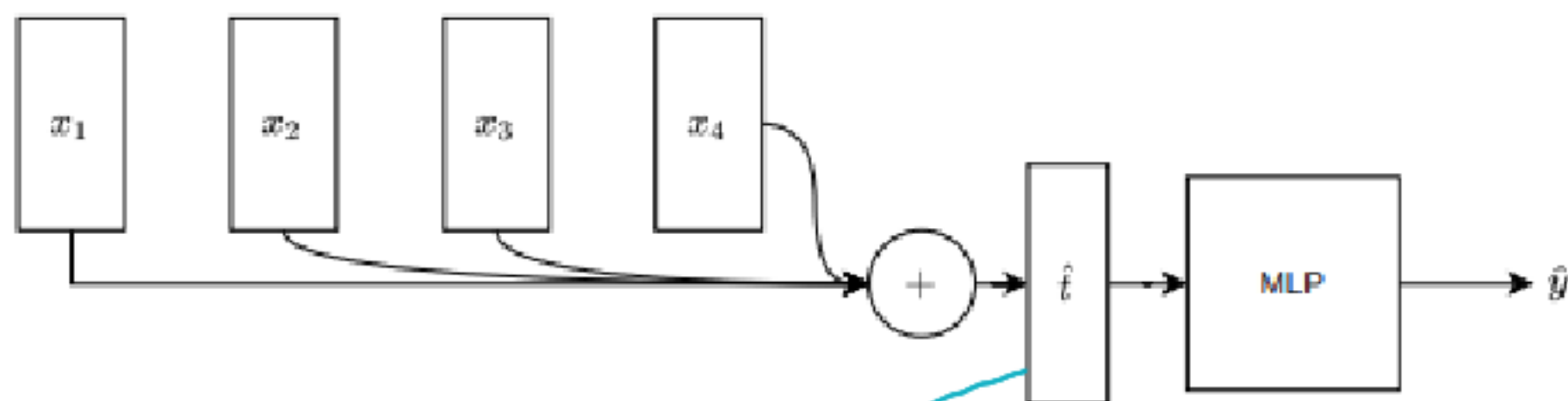
$kg$

$kg$

$cte + q \cdot x_2$

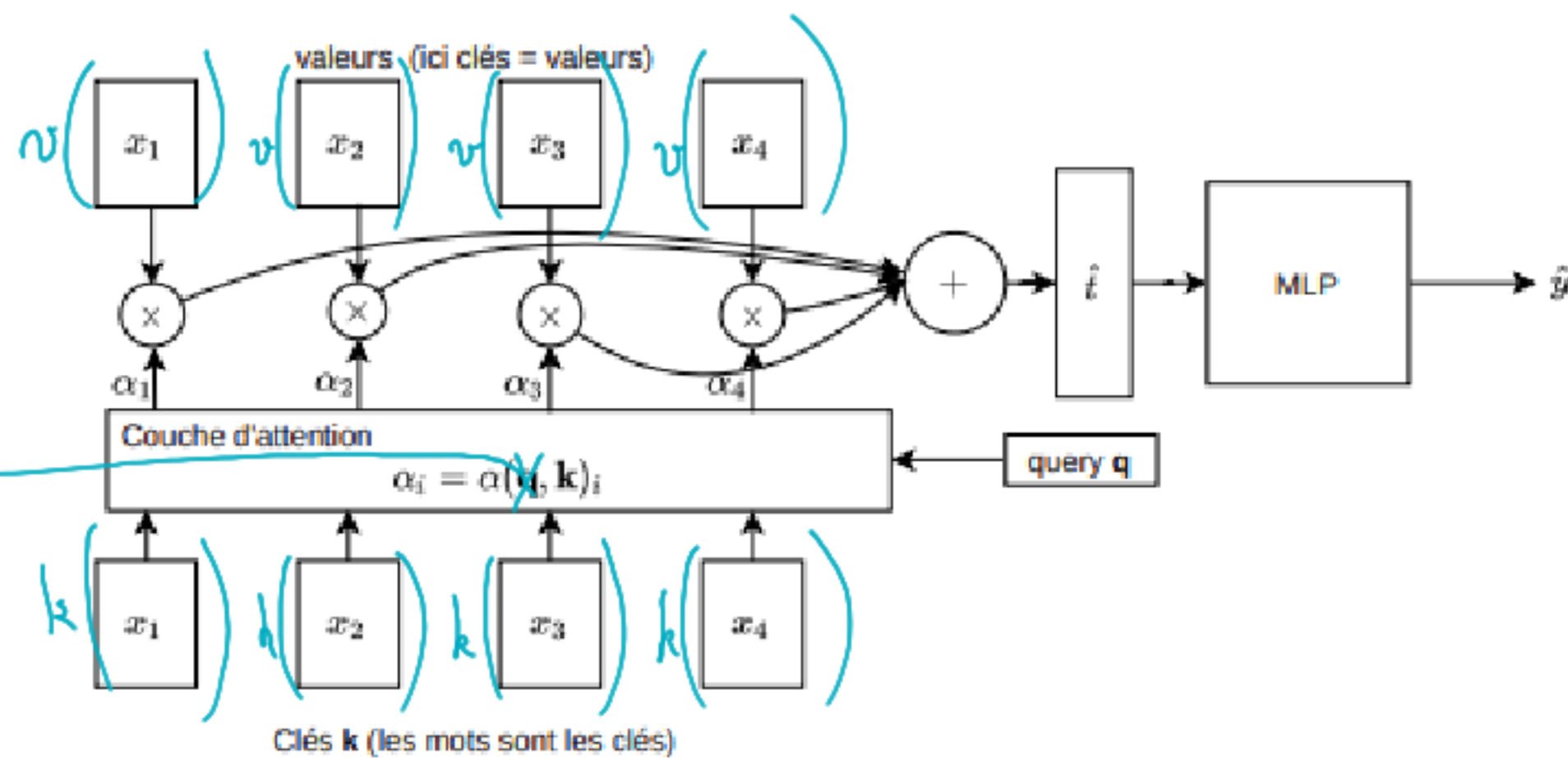
quartier  
rep. apprix

Q3



au minimum  
option  
 $q, v$  (et  $k$ )  
sont des MLP  
(appuis)

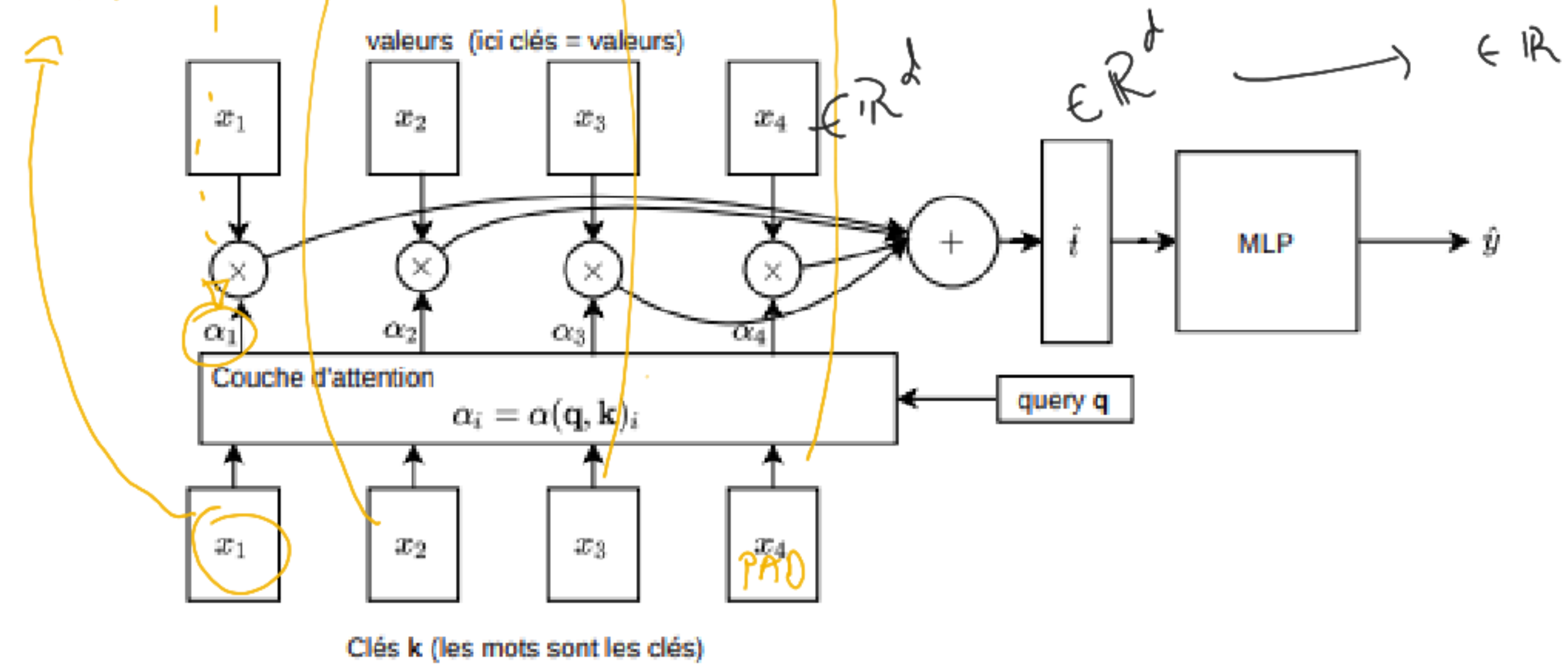
$q = \text{MLP}$   
 $q(t)$



Q2

$\alpha_1 = p(a_1 | k)$   
 $\alpha_2 = \dots$   
 $\alpha_3 = \dots$   
 $\alpha_4 = \dots$   
 $\alpha_i \cdot q$   
 $\alpha_1 \cdot q$   
 $\alpha_2 \cdot q$   
 $\alpha_3 \cdot q$   
 $\alpha_4 \cdot q$   
 $-\infty$   
softmax  
masque

ici la phrase compte 3 mots



# Mécanisme d'attention

$$\alpha = \begin{pmatrix} \alpha_{11} & \dots & \alpha_{1m} \\ \vdots & & \vdots \\ \alpha_{b1} & \dots & \alpha_{bm} \end{pmatrix} \in \mathbb{R}^{b \times m}$$

$$t \in \mathbb{N}^{b \times m} \xrightarrow{\text{glove (emb.)}} x \in \mathbb{R}^{b \times m \times d} \xrightarrow{\text{y}} y \in \mathbb{R}^{b \times m \times d}$$

$$y[i, j] = x[i, j] * \alpha_{ij} \in \mathbb{R}^d$$

$$\alpha \longrightarrow \alpha' \in \mathbb{R}^{b \times m \times 1}$$

broadcast  
 $\alpha'[i, j, k] = \alpha[i, j] \quad \forall k$

$$x \in \mathbb{R}^{b \times m \times d} \xrightarrow{\alpha'} \alpha' * x$$

↑  
 terme à terme

$$z \in \mathbb{R}^{b \times d}$$

$$z[i] = \sum_j y[i, j] = \sum_j \alpha_{ij} x[i, j]$$