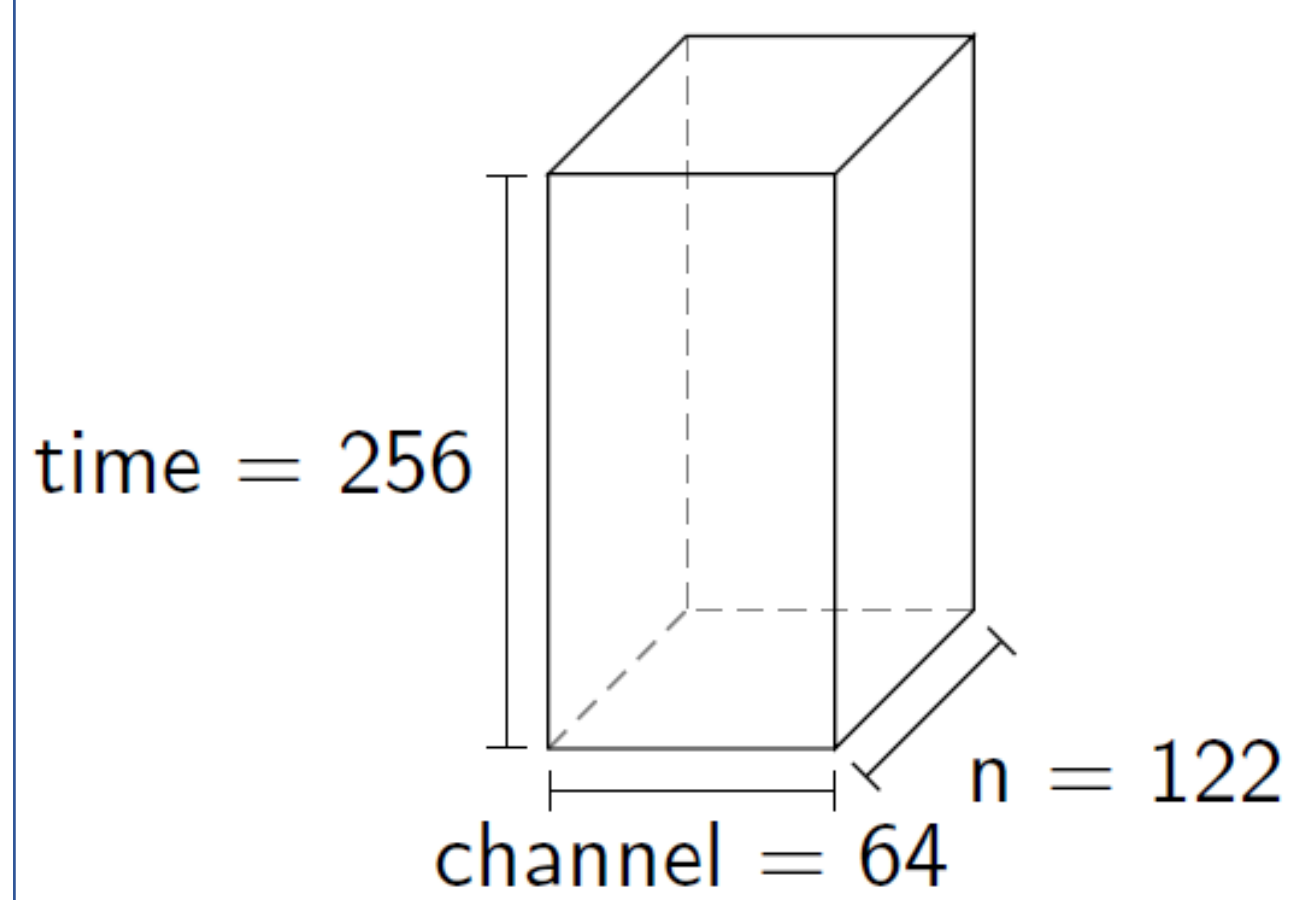


## Introduction

PCA and logistic regression has been widely used in the statistic field. However, when applying on a tensor data, MV-logistic regression has been proofed to having a better performance then the conventional logistic regression. m-PCA is also a method specially developed for the dimension reduction on tensor data.

## Data



$X_{i,j}$  = voltage value at time point  $i$  and channel of electrode  $j$   
 $Y_i$  = binary random variable with value 1 indicating alcoholism and 0 otherwise

Aim to examine EEG correlates of genetic predisposition to alcoholism.

## Matrix Variate logistic regression

$$\text{logit } P(Y = 1|X) = \gamma + \text{vec}(\eta)^T \text{vec}(M) = \gamma + A^T X B$$

with  $\eta = AB^T, A \in R^p, B \in R^q$

The parameters inquired would be  $1 + p + q$ , which is less than  $1 + pq$ .

## PCA & m-PCA

$$\text{PCA} : \text{vec}(X - \mu) = \Gamma v + \text{vec}(\epsilon)$$

$$\text{m-PCA} : \text{vec}(X - \mu) = (B_0 \otimes A_0) \text{vec}(U) + \text{vec}(\epsilon),$$

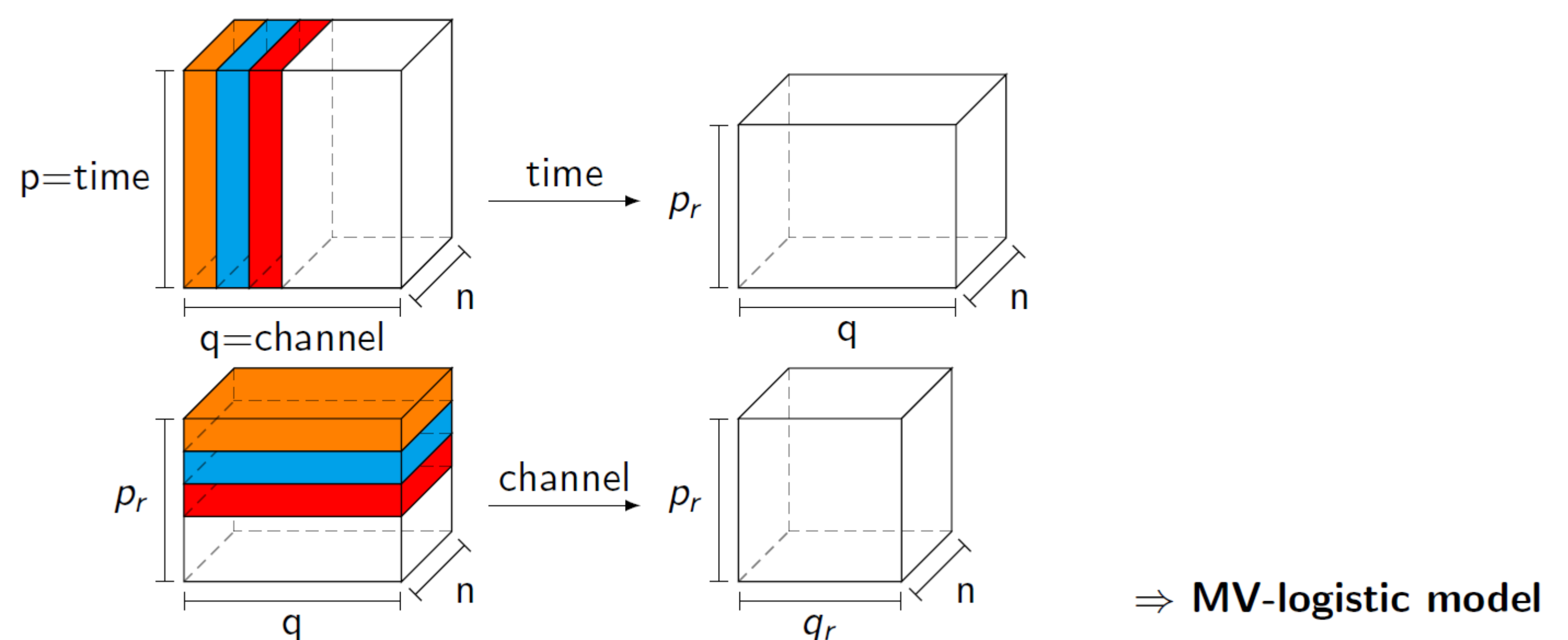
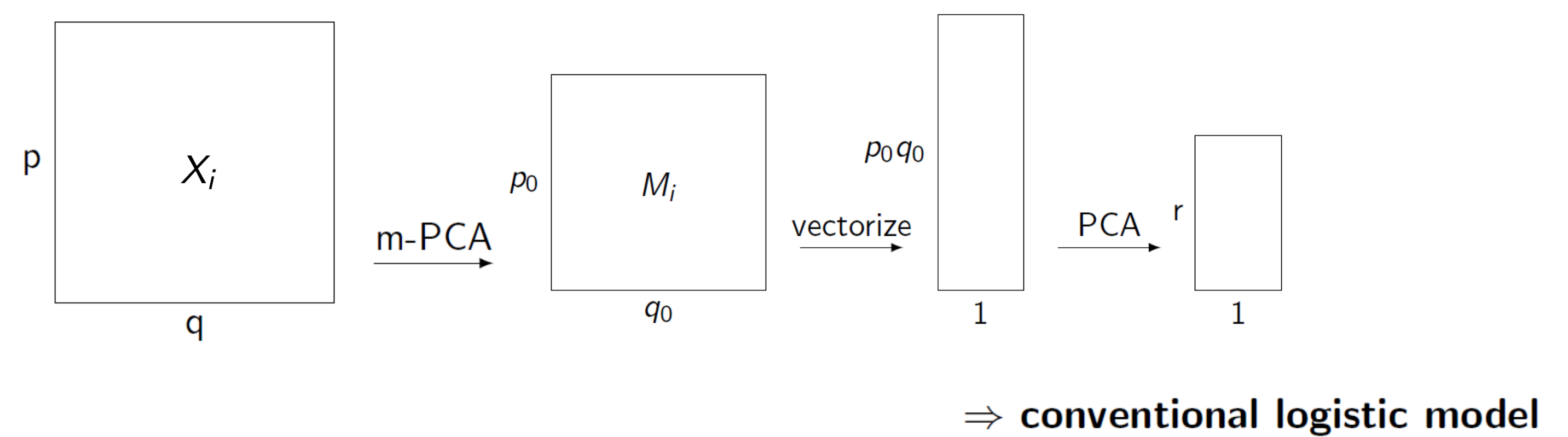
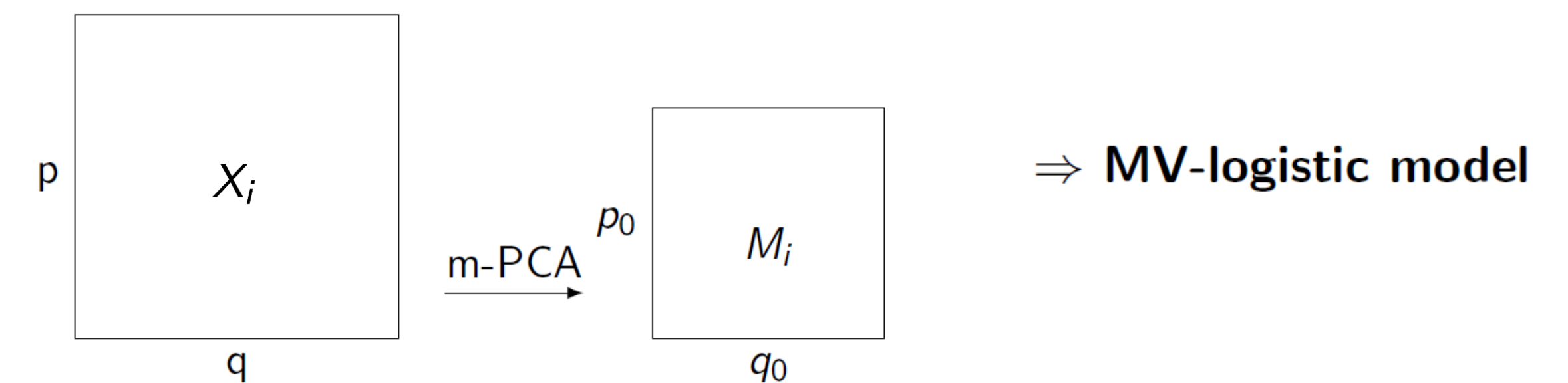
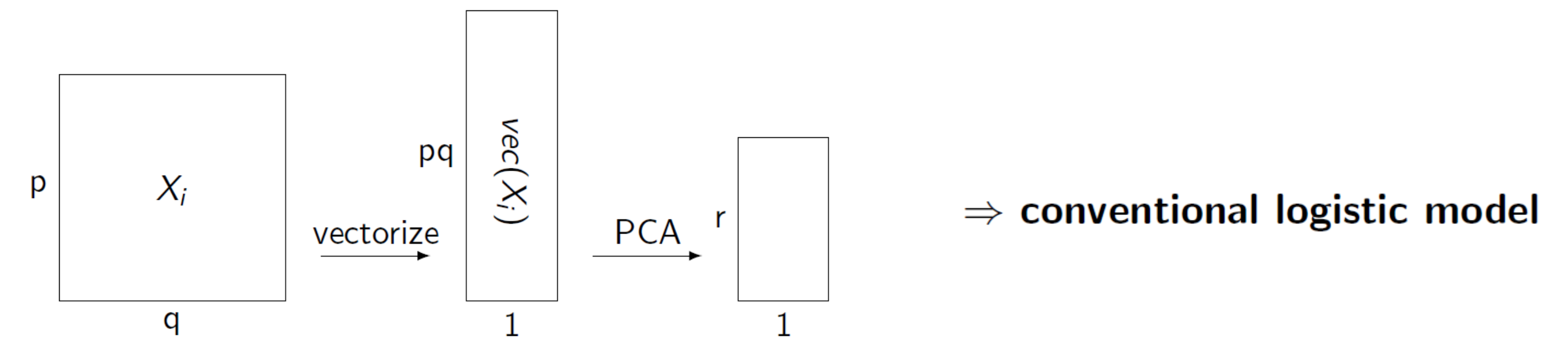
$$A_0 \in R^{p \times p_0}, B_0 \in R^{q \times q_0}$$

By leading the constraint,  $\Gamma \triangleq A \otimes B$ , make the parameters required reduced to  $2r(p + q)$ , which is less than  $1 + pq$

## Problem

1. Find a better preprocess procedure which lead to a better performance.
2. Compare m-PCA and MV,-logistic regression to the conventional models.

## Preprocessing



## Result

MV-logistic model	rank-r	2	3	4	5	6
Accuracy	0.655	0.754	0.770	0.770	0.778	
Accuracy	rank-r	7	8	9	10	20
0.803	Accuracy	0.77	0.762	0.762	0.754	0.762

Table 1. The leave-one-out classification accuracy of MV-logistic regression and the leave-one-out classification accuracies of PCA followed by conventional logistic model under different  $r$

$p_0$	$q_0$		
	15	20	30
15	0.795/0.795	0.779/0.803	0.787/0.77
20	0.811/0.820	0.746/0.754	0.730/0.762
30	0.754/0.828	0.730/0.746	0.746/0.721

Table 2. The leave-one-out classification accuracies of m-PCA followed by MV-logistic regression / PCA and conventional logistic regression, under different  $(p_0, q_0)$

$p_0$	$q_0$		
	15	20	30
15	0.795/0.828	0.779/0.770	0.787/0.746
20	0.811/0.779	0.746/0.746	0.730/0.697
30	0.754/0.738	0.730/0.705	0.746/0.738

Table 3. The leave-one-out classification accuracies of m-PCA followed by MV-logistic regression / PCA by time then by channel followed by MV-logistic regression, under different  $(p_0, q_0)$

## Discussion and conclusion

Compare to the conventional logistic regression, MV-logistic has a better performance.(Table 1) However, after using m-PCA, the PCA followed by the conventional logistic had a better accuracy. (Table 2) Implies that m-PCA is a critical component of processing even followed by a conventional methods. In the Table 3, PCA by time then by channel aim to simulate the same affect as m-PCA does. The performance only beyond the one of the m-PCA on the smallest  $(p_0, q_0)$ .

PCA is a more complex method compare to m-PCA. On other hand, m-PCA is specialized to fit on tensor data. When processing a high dimensional tensor data, we would recommend m-PCA and MV-logistic model. Even using m-PCA only could also help the performance of conventional methods followed by.