Dimension Reduction Methods

The main goal of these Methods is simply is to get the number of dimensions p which represent the **Predictors or Features** down

$$Z_m = \sum_{j=1}^{p} \phi_{jm} X_j$$

With $M < p$

- Z_m is a **linear combination** of the original predictors p

For a linear Regression example, we can fit

$$y_j = \theta_0 + \sum_{m=1}^{M} \theta_m z_{im}$$

- θ represents the regression coefficients
- $M < p$ which means the dimension of the problem has been reduced

$$\sum_{m=1}^{M} \theta_m z_{im} = \sum_{m=1}^{M} \theta_m \sum_{j=1}^{p} \phi_{jm} x_{ij} = \sum_{m=1}^{M} \sum_{j=1}^{p} \theta_m \phi_{jm} x_{ij} = \sum_{j=1}^{p} \beta_j x_{ij}$$

Formally Dimension reduction serves to constrain the estimated coefficients $\hat{\beta}$ to be a linear combination :

$$\beta_j = \sum_{m=1}^{M} \theta_m \phi_{jm}$$

Motivation behind Dimensionality Reduction

- **Storage** : especially when dealing with large amount of data
- **Model Training time** : Reduce the amount of **predictors** will largely improve the training time
- **Interpretability** : it's not valid for all the methods but Dimensionality reduction results in **spare** and less variance models which are easier to interpret and work with **inference**