PhD week 33-PCA vs Autoencoder

Ningxin Yang, PhD student

Supervisor: Dr Truong Le; Prof. Lidija Zdravkovic

Main differences:

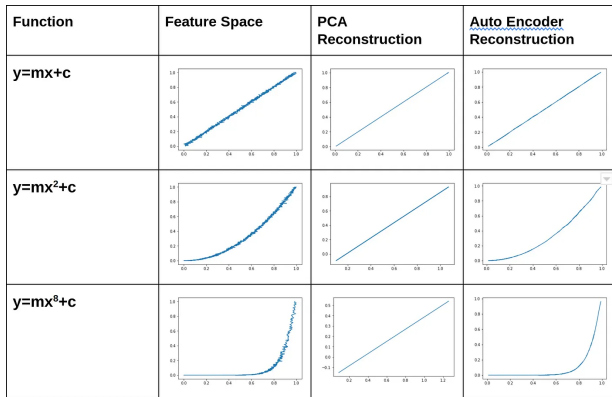- PCA is a **linear** DR technique, autoencoder is not

Function	Feature Space	PCA Reconstruction	Auto Encoder Reconstruction
y=mx+c			
y=mx²+c			
y=mx⁸+c			

Figure 1: PCA with one PC vs AE with one latent space from *UrwaMuaz*

- Syntax:

$$[coeff, score, latent, \tilde{}explained, mu] = \\ pca(data,' Economy', false) \\ PC = data * coeff(:, 1:k) \\ PCA_{reconstruct} = PC * coeff(:, 1:k)'$$

$$autoenc = \\ trainAutoencoder(data, hiddenSize) \\ AE_{PCs} = encode(autoenc, data) \\ AE_{reconstruct} = decode(autoenc, AE_{PCs})$$

Data is $n \times p$; pca doesnot require zscore

Data is $p \times n$; AE requires zscore

Determine the reduced space

Several methods exist [1, 2]:

- Akaike information criterion (AIC)
- minimum description length (MDL)
- imbedded error function (IEF)
- **cumulative percent variance (CPV)**
- **scree test on residual percent variance (RPV)**
- **average eigenvalue (AE)**
- parallel analysis (PA)
- Autocorrelation
- **cross validation**
- variance of reconstruction error

Permutation test to choose PC number

Main principle: Permutation test disorder the column arbitrarily to break the covariance matrix, personally assuming as a way of average eigenvalue.

Only feasible for linear beam deflection

Problems:

- Does orthogonal means independent? Should not
- explain variance is almost RMSE or MSE?
- The literature I read is mostly based on features with noise
- MSE percentage is not reliable (Zachary deflection)
- Cross validation, is it useful? test the covariance stable or not? Not sensitive to the observations not much (80 test 20 train)
- Why beam problem is very easy to capture the covariance?
- ICA requires we know the number of individual number
- AE not always better than PCA
- Use eye?

[1] Edoardo Saccenti and José Camacho. "Determining the number of components in principal components analysis: A comparison of statistical, crossvalidation and approximated methods". In: *Chemometrics and Intelligent Laboratory Systems* 149 (2015), pp. 99–116.

[2] Sergio Valle, Weihua Li and S Joe Qin. "Selection of the number of principal components: the variance of the reconstruction error criterion with a comparison to other methods". In: *Industrial & Engineering Chemistry Research* 38.11 (1999), pp. 4389–4401.