CS-E3210- Machine Learning Basic Principles
Home Assignment 6 - "Feature Learning"

Your solutions to the following problems should be submitted as one single pdf which does not contain any personal information (student ID or name). The only rule for the layout of your submission is that for each problem there has to be exactly one separate page containing the answer to the problem. You are welcome to use the LaTeX-file underlying this pdf, available under `https://version.aalto.fi/gitlab/junga1/MLBP2017Public`, and fill in your solutions there.

Problem 1: The Principal Component

Answer.

$$\mathcal{E}(\hat{V}, \hat{W} \mid X) = \min_{V,W \in S_d} \mathcal{E}(V, W \mid X)$$
$$= \min_{V,W \in S_d} \frac{1}{N} \sum_{i=1}^{N} \| x^{(i)} - V W^T x^{(i)} \|_2^2$$
$$= \min_{V,W \in S_d} \frac{1}{N} \| X - X W V^T \|_2^2$$

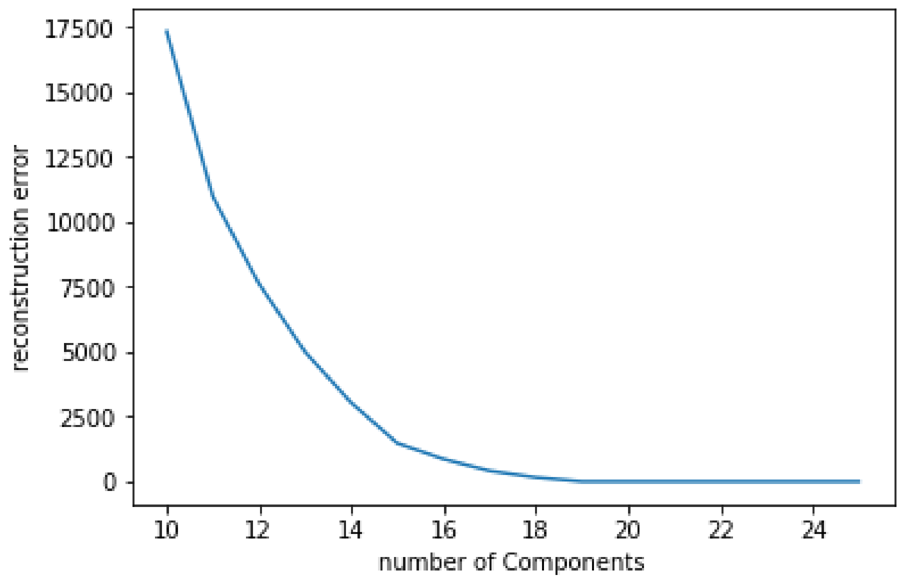
$$W V^T = \underset{\|WV^T\|=1}{\arg\max} \{ \| X W V^T \|^2 \} = \underset{\|WV^T\|=1}{\arg\max} \{ V W^T X^T X W V^T \}$$

since WV^T is a unit vector,

$$W V^T = \arg\max \left\{ \frac{V W^T X^T X W V^T}{V W^T W V^T} \right\}$$

The quantity to be maximized can be recognised as a Rayleigh quotient. A standard result for a positive semidefinite matrix such as $X^T X$ is that the quotient's maximum possible value is the largest eigenvalue of the matrix, which occur when WV^T is the corresponding eigenvector.

So WV^T is a $d \times d$ matrix whose columns are the eigenvectors of $X^T X$ ($\frac{1}{N} X^T X$)



The minimum reconstruction error is 7.9279989669496444e-23.

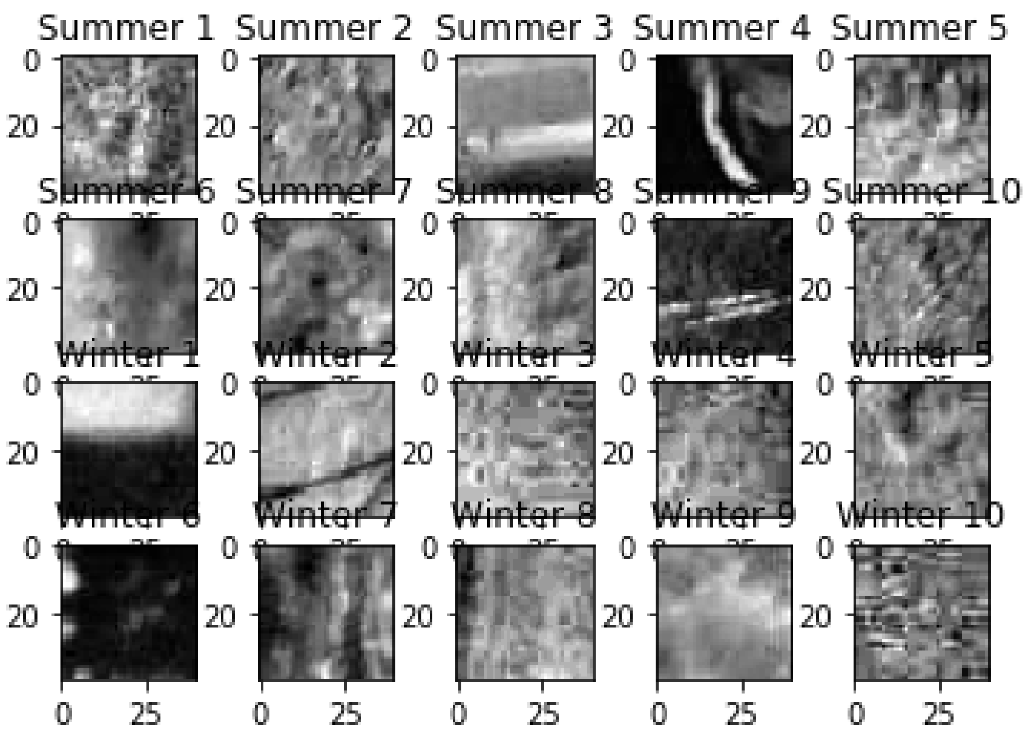


Figure 1: Reconstructed grayscale plots