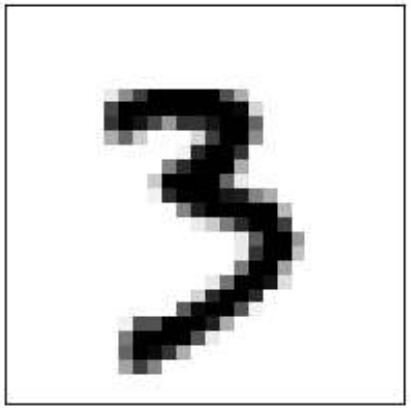
Lecture 18: Dimensionality Reduction

Professor Ilias Bilionis

Principal component analysis: Selecting the number of terms

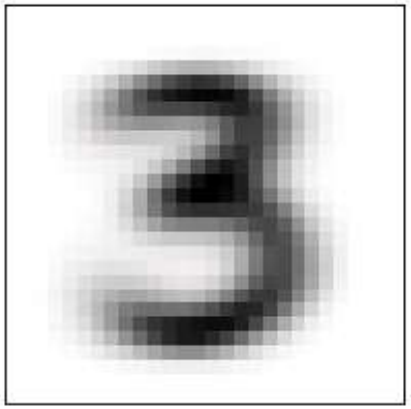
How many terms to keep?



Project

$\mathbf{z}_1, \ldots, \mathbf{z}_d$

Reconstruct



$d = 1$

$\vdots$

$d = 227$, in this case

Explained variance

$$\text{Reconstruction Error}^{(d)} = \sum_{j=d+1}^{D} \lambda_j$$

Adding the $d+1$ PC reduces rec. error by $\hat{\lambda}_{d+1}$.

$$\text{Variance of data set} = \sum_{j=1}^{D} \lambda_j \quad \text{- total}$$

$$\text{Variance explained by } d \text{ PC.} = \sum_{j=1}^{d} \lambda_j \quad \text{- up to } d$$

$$\text{Stop when } \frac{\sum_{j=1}^{d} \lambda_j}{\sum_{j=1}^{D} \lambda_j} < 0.98$$

Explained variance