EECS 16B CSM

Bryan Ngo

UC Berkeley

2020-11-15

Principal Component Analysis

Motivation

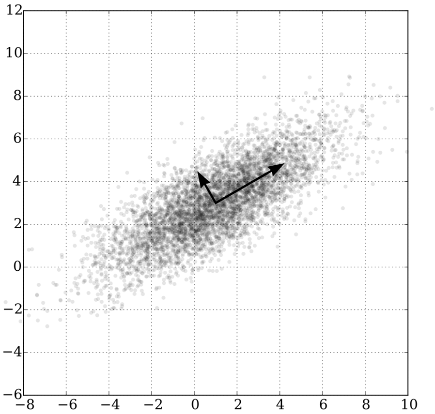
- used for statistical analysis
- clustering
- correlation

How to PCA

Given $\boldsymbol{A} \in \mathbb{R}^{n \times m}$, n measurements with m samples,

1. find $\overline{n_i}$ to center $\boldsymbol{A}$ around mean
2. find covariance matrix $\boldsymbol{C} = \frac{1}{m} \left[\widetilde{\boldsymbol{A}}\right]^\top \widetilde{\boldsymbol{A}}$
3. plot any two eigenvectors/principal components v_1, v_2 against centered points
4. data is scaled by σ_1, σ_2
5. more stretched along vector $\implies$ larger correlation

Visualization



Signals

Sampling

- continuous $\rightarrow$ discrete
- measuring an analog signal at a frequency ω
- band limiting
  - if $\omega > 2\omega_{max}$, signal perfectly recovered (Nyquist frequency)

Interpolation

- discrete $\rightarrow$ continuous
- we want to pass through every sampling point, not approximate it
- composed of a weighted sum of basis functions

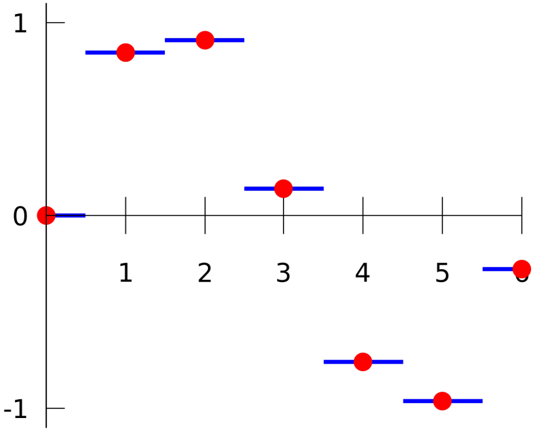
Given basis function

$$\Phi_i(x) = \begin{cases} 1 & x = i \\ 0 & \text{elsewhere} \end{cases} \tag{1}$$

$$y(x) = \sum_{i=1}^{n} y_i \Phi_i(x) \tag{2}$$

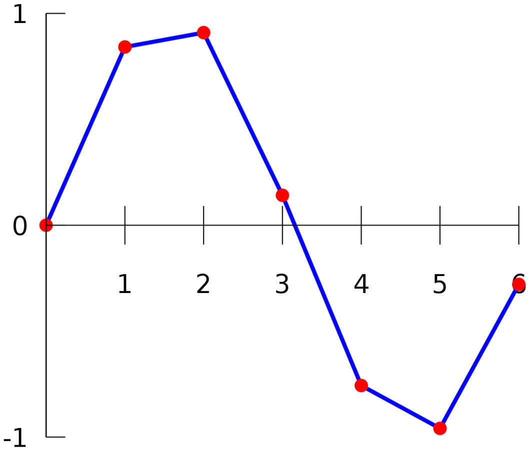
Interpolation

Zero-Hold



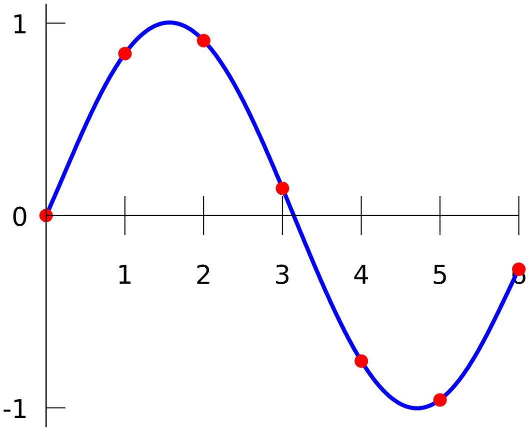
Interpolation

Linear



Interpolation

Polynomial



Interpolation
Sinc