Beyond PCA
Continuous Latent Variables
Ron Parr
CPS 271

PCA Overview
• PCA finds a linear subspace with the lowest reconstruction error
• Equivalently: PCA finds the linear subspace in which the data have the highest variance
• Used for: De-noising, feature extraction, compression, classification, etc.
• Limitations: Does not respect class labels, requires data aligned as vectors, may not return meaningful principle components, doesn’t handle missing data

Probabilistic PCA
• Suppose data are N dimensions, and we want to find d principle components
• Assume data are generated from an underlying (latent) d-dimensional vector z, which has 0 mean, Gaussian distribution
• We observe view x=Wz + u + e
  – W is an nxd matrix that expands z into N space
  – u is a N-vector offset (mean in N space)
  – e is 0 mean N-dimensional Gaussian noise
• Compare with PCA assumptions:
  – Both methods assume data come from d dimensional space
  – Both can be viewed as assuming extension in N-dimensional space occurs via linear transformation + Noise

Solving Probabilistic PCA
• The ML solution to probabilistic PCA is equivalent to the regular PCA solution (similar them to probabilistic regression)
• So, why bother?
  – Output of P-PCA is a distribution
  – We can draw samples from this distribution
  – We can estimate probabilities of new data points
  – We can handle missing data more gracefully
  – We can also do the Bayesian thing

Kernel PCA
• Keeping with current fashion, people have produced kernelized versions of PCA
• Not as clean and easy as kernelized SVMs (requires more computation and approximations)
• Think before you try this: Why do you want to expand your data into a higher dimensional space, and then find a lower dimensional linear subspace within the higher dimensional space? (There are reasons to do this, but they’re less obvious than for SVMs.)

Independent Component Analysis (ICA)
• PCA is not particularly helpful for finding independent clusters
• ICA idea:
  – Assume non-Gaussian data
  – Find multiple sets of components
  – Minimize correlation between components
• Blind source separation example:
  – Given: Audio recording with w/2 overlapping voices
  – Goal: Separate voices into separate tracks
Other Dimensionality Reduction Methods

- PCA minimizes L2 reconstruction error
- Multidimensional scaling (MDS) tries to preserve pairwise distances in the projected space
- Isometric feature mapping (ISOMAP) tries to do MDS with geodesic distances
- Locally linear embedding (LLE) tries to preserve local relationships over a set of points in the “neighborhood” of each other

Summary

- “Continuous Latent Variable” techniques assume that high dimensional data are well summarized by lower dimensional representations

  Main differences between techniques
  - Modeling assumptions
  - Optimization criterion