ECE 592
Topics in Data Science

Dror Baron
Associate Professor
Dept. of Electrical and Computer Engr.
North Carolina State University, NC, USA
Dimensionality Reduction

Keywords: principal components analysis, random projections
What is Dimensionality Reduction?
What is it?

- Wikipedia:

In machine learning and statistics, *dimensionality reduction* or dimension reduction is the process of reducing the number of random variables under consideration, via obtaining a set of principal variables. It can be divided into feature selection and feature extraction.
Principal Components Analysis

[Bishop, Chapter 12.1]

Keywords: principal components analysis
What is PCA?

- Have point cloud of data
- Can apply transformation

- Want transformation that arranges most of variance in point cloud among few dimensions
- Might be able to throw away dimensions with little information
- *We’ll soon compute this transform*
Application – identifying digits

- Want to identify digits

- Apply PCA to each digit
  - Identify $M$ principal components, $u_1, \ldots, u_M$
  - Multiply data by principal components
  - Large numbers $\rightarrow$ digit probably appeared; else it’s unlikely (hypothesis testing)
Contrast with sparsifying transform

- Sounds similar to sparsifying transform; energy compaction

- Sparsifying transform – we know the signal is sparse; we don’t (necessarily) know where big coeffs are
  - Emphasis is “given the support set, it’s very sparse”

- PCA – we (statistically) know where the big coeffs are
  - Emphasis is “we know where the large coeffs are”
Derivation

- Point cloud comprised of N vectors, \( x_n \in \mathbb{R}^D \)
- Begin with *best* projection vector, \( u_1 \in \mathbb{R}^D \)
  - It packs as much average energy as possible

- **Expectation**: \( E[u_1^T x] = u_1^T E[x] \)

- **Variance**: \( \text{var}(u_1^T x) = E[(u_1^T x - E[u_1^T x])^2] \)
- Can show \( \text{var}(u_1^T x) = u_1^T S u_1 \)
- S covariance matrix of data,\[ S = \frac{1}{N} \sum_{n=1}^{N} (x_n - E[x]) (x_n - E[x])^T \]
Derivation continued

- Want $u_1$ with largest $\text{var}(u_1^T x) = u_1^T S u_1$
- Add unit norm constraint, $||u_1||=1$
- Solve $u_1$ by maximizing $u_1^T S u_1 + \lambda (1 - u_1^T u_1)$
  - Lagrangian $\lambda$

- Can show that solution satisfies $S u_1 = \lambda_1 u_1$
  - Implies that $(\lambda_1, u_1)$ are eigen-(value,vector) pair of $S$
  - Variance satisfies $u_1^T S u_1 = u_1^T \lambda_1 u_1 = \lambda_1$

- Optimal M-dimensional projection corresponds to M eigen-vectors corresponding to largest eigen-values
Random Projections

Keywords: random projections
Why random?

- We may not be able to compute principal components
- Or PCA may suffer from outliers, etc.

**Solution:** compute $M < D$ random projections of data

- On average, especially in high dimensions, random projections are approximately energy preserving
  - Preserves geometric structure of space
  - Surprisingly, random projections are not far from optimal