Dimensionality Reduction using PCA I

Overview:

- Let d be the number of features in each datapoint.
- Dimensionality Reduction: Find fewer features (less than d) such that they are representative of the original d features.
- Principal Component Analysis (PCA) gives us new features (principal components) that are linear combinations of the original features.
- Principal components can be sorted by the amount of variance in the data that they can explain.
- Principal components form an orthonormal basis of the d dimensional feature space.
- p principal components ($p < d$) are chosen such that maximum variance in the data is captured.

Dimensionality Reduction using PCA II

- Let $\mathbf{X}$ be an $n \times d$ data matrix that contains n datapoints having d feature dimensions.
- Mean centering: The average value of each of the d features is subtracted from the rows of $\mathbf{X}$. This produces a dataset such that the mean of the rows is $\vec{0}$.
- We will assume that the data matrix $\mathbf{X}$ is mean centered.
- Covariance between features f_1 and f_2:
  $COV(f_1, f_2) = \frac{1}{n-1} \sum_{i=1}^{n} (f_1 - \bar{f_1})(f_2 - \bar{f_2})$
- Step 1: Find covariance matrix $cov = \frac{1}{n-1} \mathbf{X}^{\mathbf{T}} \mathbf{X}$
  Note: If the scales of the features vary widely, then compute the correlation matrix instead of covariance matrix in Step 1.
- The $(i, j)^{th}$ element of cov matrix will contain the covariance between i^{th} and j^{th} features.

Dimensionality Reduction using PCA III

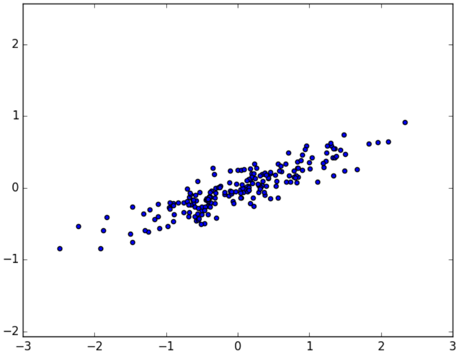
- $rank(cov) = rank(\mathbf{X}) = r$ (because $rank(\mathbf{X}^\mathbf{T}\mathbf{X}) = rank(\mathbf{X})$).
- eigenvectors of cov will be orthogonal (because cov is symmetric).
- Step 2: Find the eigenvalues and orthonormal eigenvectors of cov. Order the eigenvectors by their eigenvalues from highest to lowest.
- The eigenvectors corresponding to large eigenvalues are more significant, and capture more variation in the data.
- Step 3: Select p eigenvectors (principal components) corresponding to the p largest eigenvalues.

Dimensionality Reduction using PCA IV

- Step 4: Construct a matrix $\mathbf{V}$ with p orthonormal eigenvectors as its columns.
- Step 5: Project the data matrix $\mathbf{X}$ on to the space spanned by the p eigenvectors in $\mathbf{V}$.
  $\mathbf{X_{new}} = \mathbf{XV}$
- $\mathbf{X_{new}}$ is the new data matrix with each datapoint having p dimensions.

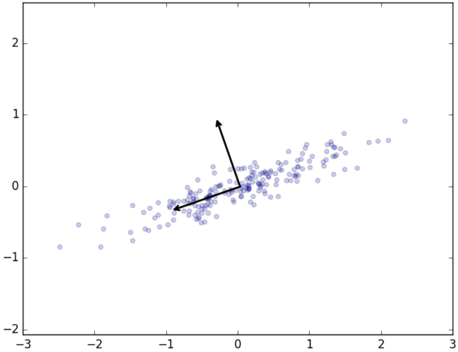
Dimensionality Reduction: An Example I

- Let $\mathbf{X}$ be mean centered 200×2 datamatrix. Each datapoint has two dimensions.



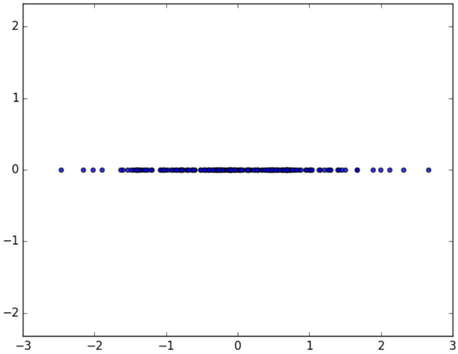
Dimensionality Reduction: An Example II

- Figure shows the orthonormal eigenvectors of *cov* matrix.



Dimensionality Reduction: An Example III

- Figure shows vectors in $\mathbf{X_p} = \mathbf{XV}$, where $\mathbf{V}$ is 2×1 matrix containing the eigenvector corresponding to the largest eigenvalue. The dimensionality is reduced while retaining maximum possible variation in the data.



Dimensionality Reduction: An Example IV

- The reduced data $\mathbf{X_p}$ can be moved back to the original two dimensional feature space using $\mathbf{X_p V^T}$. The figure shows datapoints in $\mathbf{X_p V^T}$ as dark dots, and those in the original data matrix $\mathbf{X}$ as light dots.