# PCA Properties

Data Mining
Prof. Dawn Woodard
School of ORIE
Cornell University

# Outline

# Principal Components

**Property 1: PCs are only unique up to a factor of $-1$.**

- We get the principal components by singular value decomposition: $X = UDV^T$. The columns of $V$ are the orthogonal eigenvectors of $\mathrm{Cov}(X) = \frac{1}{n-1}X^TX$ that have norm 1.

- But these are only unique up to a factor of -1! Why?

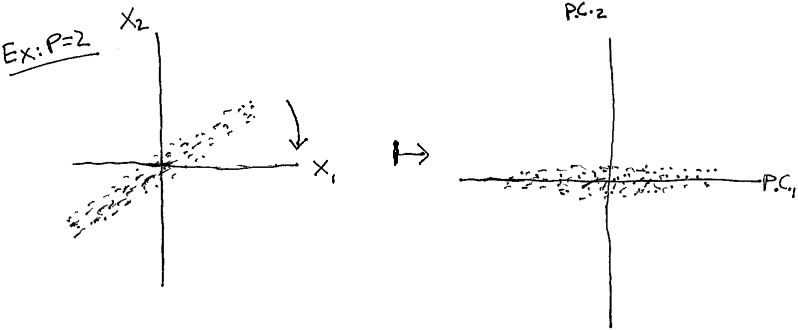
- Practically speaking, this means that if you run PCA twice:

# Principal Components

**Property 2: PCA is a rotation of the variables in $\mathbb{R}^p$.**

- PCA is a rotation of the data in $\mathbb{R}^p$, from $X$ to $XV$.
- It's a rotation because $V$ is an orthogonal matrix, $V^T V = I_p$.
- (Strictly speaking, it is a rotation possibly with a reflection, because of the invariance from the previous slide).

# Principal Components




Remember that P.C.$_j$ means the $j$th column of $XV = UD$.

Notice that although $X_1$ and $X_2$ are correlated, P.C.$_1$ and P.C.$_2$ are not.

# Principal Components

**Property 3: Principal components are uncorrelated, and are ordered by their variances.**

Proof:

# Principal Components

- I.e., the first P.C. is the linear combination of the original variables $X_1, \ldots, X_p$ that has the highest variance
- The second P.C. is:
- The third P.C. is:
- and so on.

# Principal Components

**So we can potentially capture much of the information in the *p* original variables with just the first few P.C.s! This is how PCA is used for dimension reduction.**

# Principal Components

**Property 4: Can measure how much "information" is in the $j$th PC by $\frac{d_j^2}{\sum_{k=1}^p d_k^2}$.**

- **Claim:** "total sample variance is preserved under rotation," i.e.

$$\text{Var}(X_1) + \ldots + \text{Var}(X_p) = \text{Var}(\text{P.C.}_1) + \ldots + \text{Var}(\text{P.C.}_p)$$

  (will prove)

- So we can measure "how much information is in the $j$th P.C." using the % of total variance that the $j$th P.C. has:

$$\frac{\text{Var}(\text{P.C.}_j)}{\text{Var}(\text{P.C.}_1) + \ldots + \text{Var}(\text{P.C.}_p)} = \frac{d_j^2}{\sum_{k=1}^p d_k^2}$$

# Principal Components

- Also remember that $P.C._1$ has the highest variance, $P.C._2$ has the next-highest variance, etc.

- So we can look at the proportion of total variance contained in the first $q$ P.C.s as a measure of "how much of the information from the original data is captured in the first $q$ P.C.s"

$$\frac{\mathrm{Var}(P.C._1) + \ldots + \mathrm{Var}(P.C._q)}{\mathrm{Var}(P.C._1) + \ldots + \mathrm{Var}(P.C._p)}$$

- This can help us decide how to choose $q$, i.e. how much dimension reduction can we do without losing too much of the original information (e.g. may want to keep at least 90% of original info.).

# Principal Components

**Proof:** that “total variation is preserved under rotation,” i.e. under multiplication by an orthogonal matrix $V$:

$$\mathrm{Cov}(X) = \frac{1}{n-1} X^T X$$

$$\mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_p) = \sum_{j=1}^{p} \mathrm{Cov}(X)_{jj} = \mathrm{trace}(\mathrm{Cov}(X))$$

$$\mathrm{Cov}(XV) = \frac{1}{n-1} V^T X^T X V$$

$$\text{sum of variances} = \sum_{j=1}^{p} \mathrm{Cov}(XV)_{jj} = \mathrm{trace}(\mathrm{Cov}(XV))$$

# Principal Components

Property of trace: trace($AB$) = trace($BA$) where $A$, $B$ are matrices. So

# Principal Components

**Example 1:**

- Say we have a dataset $X$ with $p = 2$ and

$$\mathrm{Cov}(X) = \frac{1}{n-1} X^T X = \begin{pmatrix} 1 & .7 \\ .7 & 1 \end{pmatrix}$$

- What are the eigenvectors and eigenvalues of $\mathrm{Cov}(X)$? (Guess-and-check)

# Principal Components

**Example 1:**

- The columns of $V$ are the eigenvectors of Cov($X$) having norm = 1, so $V =$
- Notice that this is also correct if we multiply any of its columns by $-1$.
- The variance of the first P.C. is:
- The variance of the 2nd P.C. is:

# Principal Components

**Example 1:**

- If the first observation in the dataset has $x_{i1} = -3$ and $x_{i2} = 2.2$, what are the values of the P.C.s for this observation?

- What % of the information in the data is captured by the first P.C.?

# Principal Components

**Example 2: What happens if the original variables are uncorrelated?**

- Then the empirical covariance matrix $\frac{1}{n-1}X^TX$ is close to diagonal.
- Assuming the variables have been standardized and taking e.g. $p = 2$, $\text{Cov}(X) \approx \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.
- This has eigenvalue 1 of multiplicity 2, i.e. $\frac{d_1^2}{n-1} = \frac{d_2^2}{n-1} = 1$.
- The eigenvectors of $\text{Cov}(X)$ are nonunique–any vector is an eigenvector with eigenvalue 1!
- So the P.C.s are nonunique. It is not useful to apply PCA in this case.