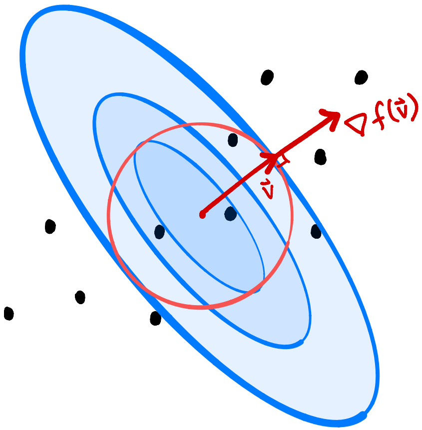


CSE 151A

Intro to Machine Learning

Lecture 18 – Part 01

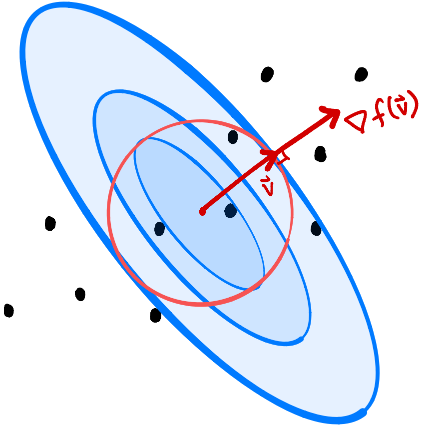
Ending the Quarter

Final Exam

- Section of Final Exam covering Weeks 09 and 10 is cancelled.
- (Optional) Redemption Sections will still be given.

Grades

- Midterm 02 grades will be posted tonight or tomorrow.
- All HWs (except for HW 08) will be posted soon.

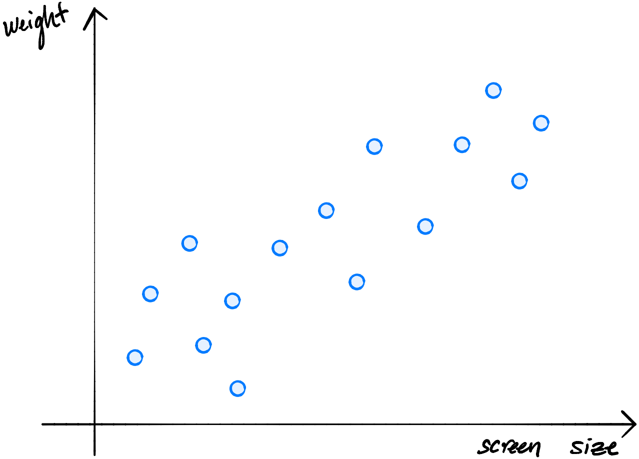


CSE 151A

Intro to Machine Learning

Lecture 18 – Part 02

PCA in Many Dimensions


weight
screen size

Recap: Principal Components

- **Goal**: Find unit vector $\vec{u}$ maximizing $\vec{u}^T C \vec{u}$.
  - I.e., find unit vector in direction of maximum variance.

- Any solution must satisfy $2C\vec{u} = \lambda\vec{u}$
  - I.e., it must be an eigenvector of C.

- The **top eigenvector** of the covariance matrix points in direction of maximum variance.

Principal Component Analysis

- **Given**: data $\vec{x}^{(1)}, \dots, \vec{x}^{(n)} \in \mathbb{R}^d$
- **Map**: each data point $\vec{x}^{(i)}$ to a single feature, z_i.
- **PCA**: Let $z_i = \vec{x}^{(i)} \cdot \vec{u}$, where $\vec{u}$ is top eigenvector of covariance matrix.

PCA in Many Dimensions

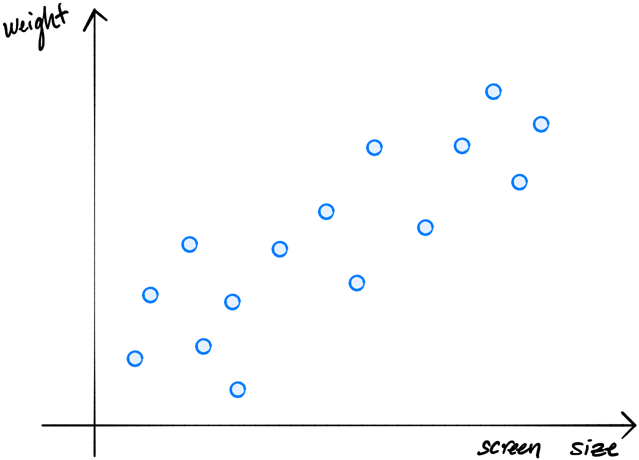
- We lose a lot of information when we project from $\mathbb{R}^d$ to a single number.

- Instead of a single number, represent $\vec{x}^{(i)} \in \mathbb{R}^d$ as a smaller vector, $\vec{z}^{(i)} \in \mathbb{R}^{d'}$.

PCA in Many Dimensions

- First PCA feature: project $\vec{x}^{(i)}$ onto principal component, $\vec{u}^{(1)}$
  - $\vec{u}^{(1)}$ is vector maximizing $\vec{u}^T C \vec{u}$.
- Second PCA feature: project $\vec{x}^{(i)}$ onto ???
- Third PCA feature: project $\vec{x}^{(i)}$ onto ???
- ...

PCA in Many Dimensions



PCA in Many Dimensions

- Second PCA feature: project $\vec{x}^{(i)}$ onto $\vec{u}^{(2)}$.
  - $\vec{u}^{(2)}$ is the **second principal component**
  - Found by maximizing $\vec{u}^T C \vec{u}$ subject to constraint that $\vec{u}^{(2)}$ is orthogonal to $\vec{u}^{(1)}$

PCA in Many Dimensions

- Third PCA feature: project $\vec{x}^{(i)}$ onto $\vec{u}^{(3)}$.
  - $\vec{u}^{(3)}$ is the **third principal component**
  - Found by maximizing $\vec{u}^T C \vec{u}$ subject to constraint that $\vec{u}^{(3)}$ is orthogonal to $\vec{u}^{(1)}$ and $\vec{u}^{(2)}$

PCA in Many Dimensions

- **Goal**: find k unit vectors $\vec{u}^{(1)}, \dots, \vec{u}^{(k)}$ such that:
  - $\left(\vec{u}^{(1)}\right)^T C \vec{u}^{(1)}$ is maximized
  - $\left(\vec{u}^{(2)}\right)^T C \vec{u}^{(2)}$ is maximized s.t. $\vec{u}^{(2)} \perp \vec{u}^{(1)}$
  - ...
  - $\left(\vec{u}^{(k)}\right)^T C \vec{u}^{(k)}$ is maximized s.t. $\vec{u}^{(k)} \perp \vec{u}^{(1)}, \dots, \vec{u}^{(k-1)}$

The New Features

- Suppose we have found $\vec{u}^{(1)}, \dots, \vec{u}^{(k)}$.
- The new feature vector for $\vec{x}^{(i)}$ is:

$$\vec{z}^{(i)} = \left(\vec{x}^{(i)} \cdot \vec{u}^{(1)},\ \vec{x}^{(i)} \cdot \vec{u}^{(2)},\ \dots,\ \vec{x}^{(i)} \cdot \vec{u}^{(k)}\right)^T$$

The New Features

- Equivalently, define:

$$U = \begin{pmatrix} \leftarrow & \vec{u}^{(1)} & \rightarrow \\ \leftarrow & \vec{u}^{(2)} & \rightarrow \\ & \dots & \\ \leftarrow & \vec{u}^{(k)} & \rightarrow \end{pmatrix}$$

- Then $\vec{z}^{(i)} = U\vec{x}^{(i)}$

The Solution

- How do we find $\vec{u}^{(1)}, \ldots, \vec{u}^{(k)}$?
- We know $\vec{u}^{(1)}$ is the top eigenvector of C.
- Claim: $\vec{u}^{(i)}$ is the ith eigenvector of C.

Spectral Theorem

- Let C be the $d \times d$ covariance matrix of X.
- In $O(d^3)$ time, we can compute its **eigendecomposition**, consisting of
  - real **eigenvalues** $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$
  - corresponding **eigenvectors** $\vec{u}^{(1)}, \ldots, \vec{u}^{(d)} \in \mathbb{R}^d$ that are orthonormal (unit length and at right angles to each other)

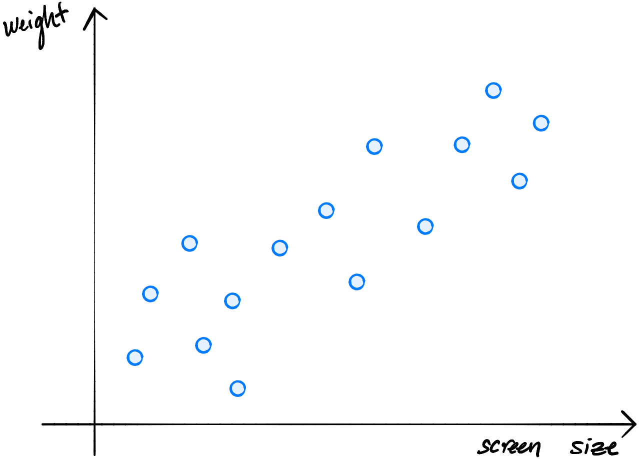
PCA

- Suppose we wish to map a data set $\vec{x}^{(1)}, \ldots, \vec{x}^{(n)}$ to k dimensions while retaining as much variance as possible.

- **Solution**:
  - Compute top k eigenvectors of covariance.
  - Place them row-wise into a matrix U.
  - $\vec{x}^{(i)} \mapsto U\vec{x}^{(i)}$.

PCA Revisited

- We have seen PCA as an optimization problem.
- Another (equivalent) view: decorrelation.

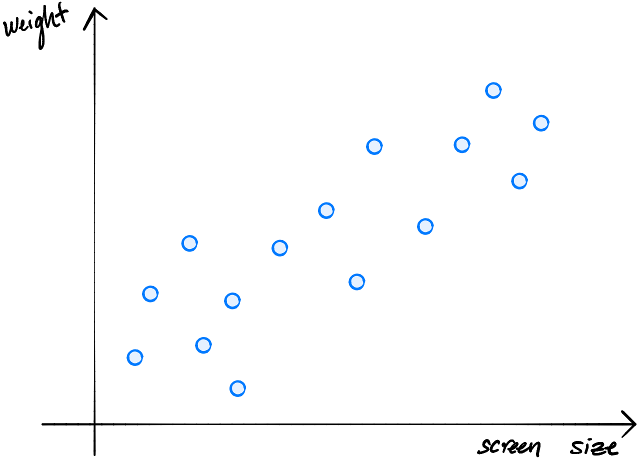

weight
screen size

Decorrelation

- **Goal**: find orthonormal basis in which data is decorrelated
  - Covariance matrix of new features is diagonal.
- **Solution**: use as basis the eigenvectors of covariance matrix.

Reconstruction

- The whole goal of PCA is to reduce dimensionality.
  - Example: turn a 784-dimensional image vector into 200-dimensional feature vector.
- But sometimes it is fun to go the other direction:
  - Take result of PCA and **reconstruct** the original image.


weight
screen size

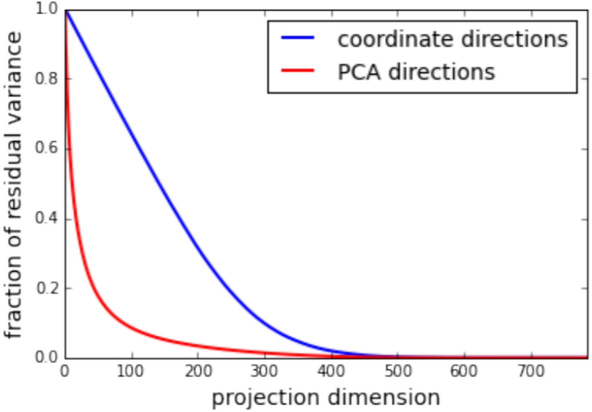
Reconstruction

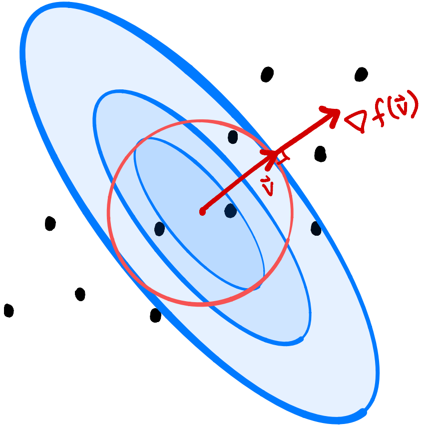
- Suppose original data is d dimensional.
- Project onto k eigenvectors $\vec{u}^{(1)}, \ldots, \vec{u}^{(k)} \in \mathbb{R}^d$.
  - $\vec{z}^{(i)} = (\vec{x}^{(i)} \cdot \vec{u}^{(1)}, \ldots, \vec{x}^{(i)} \cdot \vec{u}^{(k)})^T$
- Reconstruction of $\vec{x}^{(i)}$ from $\vec{z}^{(i)}$:

$$\vec{x}^{(i)} \approx \vec{z}_1^{(i)} \vec{u}^{(1)} + \vec{z}_2^{(i)} \vec{u}^{(2)} + \ldots + \vec{z}_k^{(i)} \vec{u}^{(k)}$$

PCA in Practice

- PCA is often used in **preprocessing** before classifier is trained, etc.
- Must choose number of dimensions, k.
- One way: cross-validation.
- Another way: the elbow method.


coordinate directions
PCA directions
fraction of residual variance
projection dimension
1.0
0.8
0.6
0.4
0.2
0.0
0
100
200
300
400
500
600
700



CSE 151A

Intro to Machine Learning

Lecture 18 – Part 04

Demos

https://go.ucsd.edu/3exnmrw