# Autoencoders

Sanjoy Dasgupta

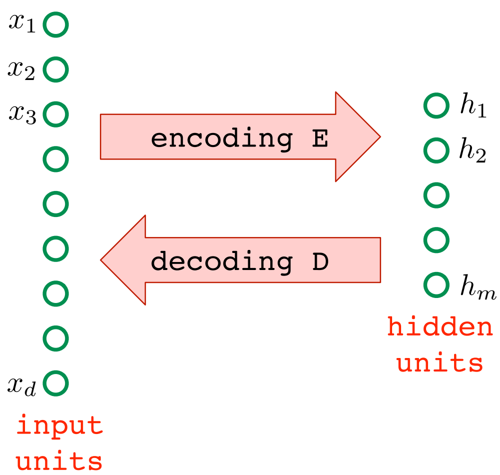
University of California, San Diego

## Topics we'll cover

1. Autoencoders
2. $k$-means and PCA as autoencoders
3. Manifold learning
4. Independent component analysis
5. Stacked autoencoders

## Autoencoders

Finding the **underlying degrees of freedom** of data

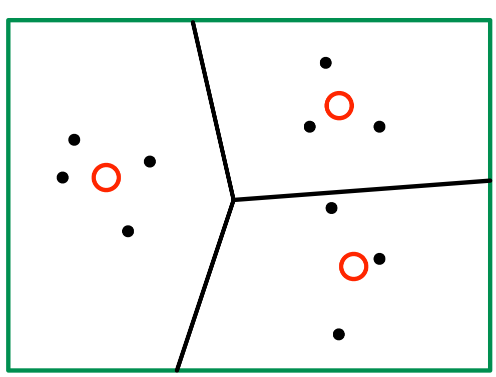



Ideally $x \approx D(E(x))$ on data points $x \in \mathbb{R}^d$

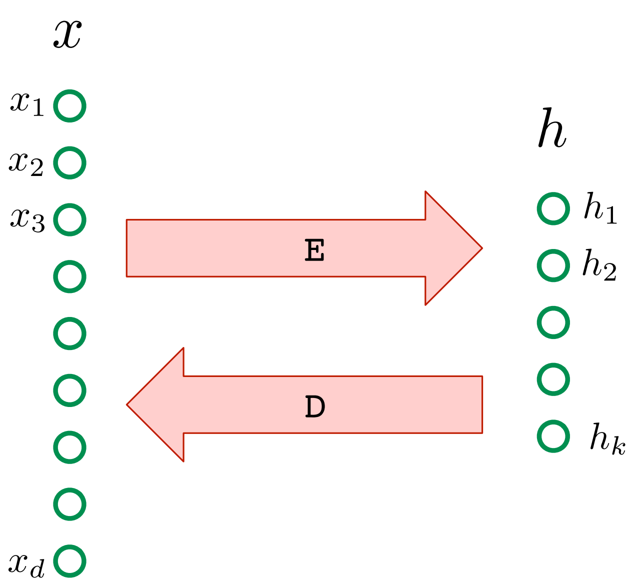
## The $k$-means clustering scheme, revisited

The $k$-means problem:

- Given: $x^{(1)}, \ldots, x^{(n)} \in \mathbb{R}^d$; integer $k$
- Find: $k$ centers $\mu_1, \ldots, \mu_k \in \mathbb{R}^d$ that minimize $\sum_{i=1}^{n} \min_{1 \leq j \leq k} \|x^{(i)} - \mu_j\|^2$



## The $k$-means autoencoder




## Principal component analysis, revisited

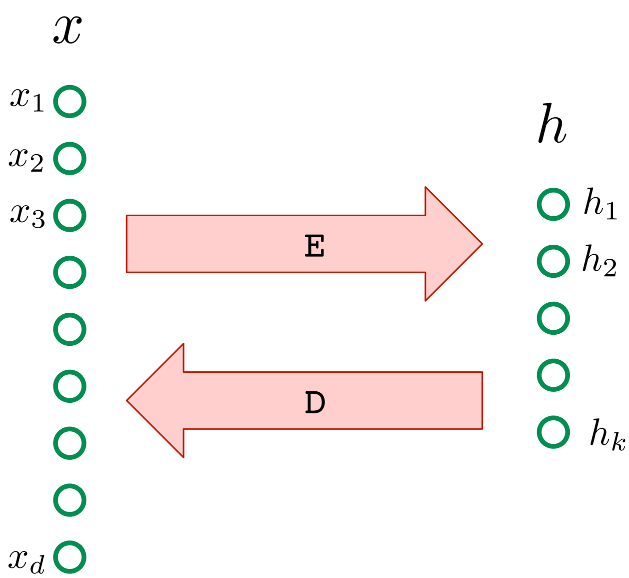
The PCA problem:

- Given: $x^{(1)}, \ldots, x^{(n)} \in \mathbb{R}^d$; integer $k$
- Find: the projection $\mathbb{R}^d \to \mathbb{R}^k$ that maximizes the variance of the projected data

Solution:

- Compute the covariance matrix of the data
- Let $u_1, \ldots, u_k$ be the top $k$ eigenvectors of this matrix
- Let $k \times d$ matrix $U$ have the $u_i$ as its columns
- Projection: $x \mapsto U^T x$
- Reconstruction: $z \mapsto Uz$

## The PCA autoencoder



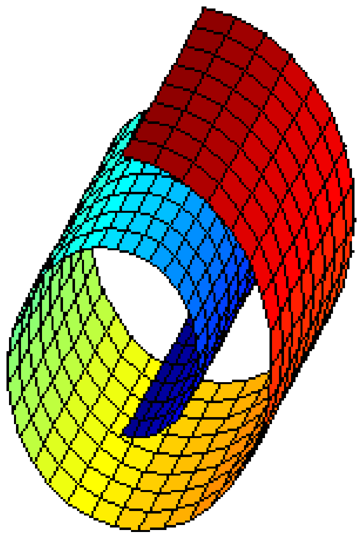

## Some other types of intrinsic structure

1. Manifold learning
   The data lies on a $k$-dimensional manifold.

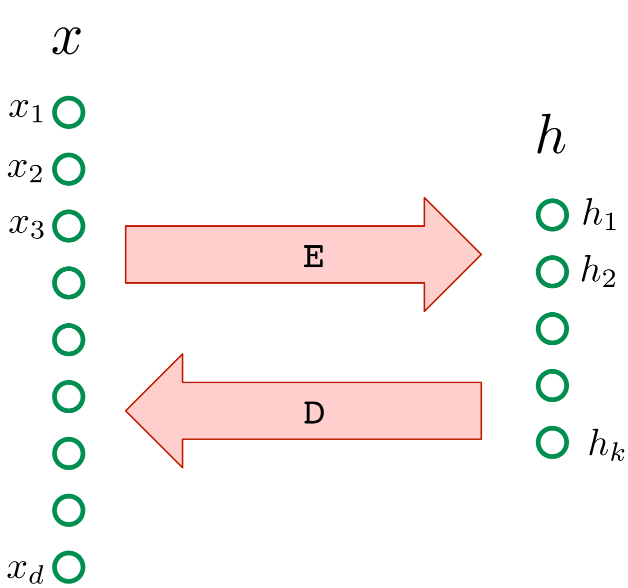
2. Independent component analysis
   The data are linear combinations of hidden features that are independent.

## Manifold learning

Sometimes data in a high-dimensional space $\mathbb{R}^d$ in fact lies close to a $k$-dimensional manifold, for $k \ll d$

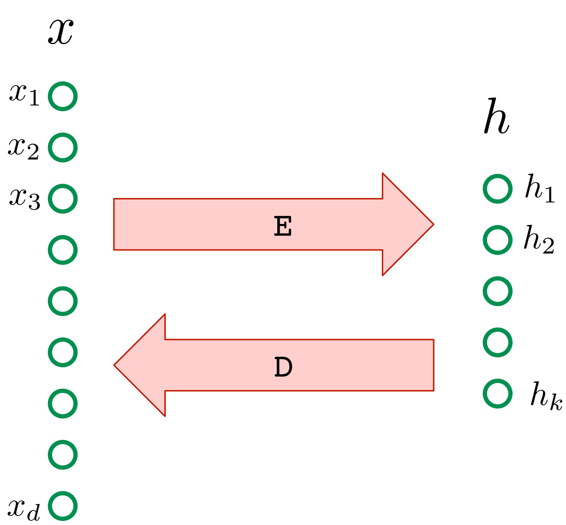


## The manifold autoencoder

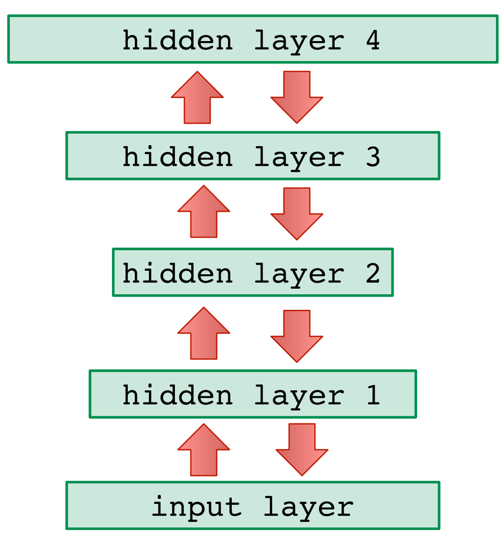


## Independent component analysis

The cocktail party problem




## Stacked autoencoders




- Fit one layer at a time to the previous layer's activations
- Then fine-tune whole structure to minimize reconstruction error