ch16: Dimensional Reduction

Definition(P5)

Reduce the number of dimensions (attributes) of the data while preserving the intrinsic characteristics of the data

在数据的所有维度中，找最能反映数据本质属性的那些

Try to identify some "hidden" aspects that determines the characteristics of the data

Why Reduce Dimensionality(P6)

- 5 reasons

Overview of Methods(P7)

- Linear Method
  - **Principal component analysis** (PCA)
  - Linear discriminate analysis (LDA)
  - Factor analysis
  - ...

- Non-linear Method
  - KPCA, KLDA
  - Multidimensional scaling (MDS)
  - t-SNE
  - **Autoencoder**
  - ...

PCA

Principal Component Analysis (PCA)

> 只关心能够最大程度反映数据变化的方向和维度

Principal component analysis seeks a space of lower dimensionality, in which the **variance** of the projected data are maximized.

– try to model how data are spread by maximizing the variance of the projected data

– find *k* orthogonal vectors that represent the "spread tendency" of the data ($k << d$) 找到k个表示数据的"扩散趋势"的正交向量

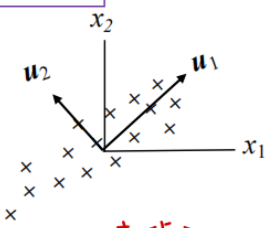
– the *k* vectors will be used as the bases of the new space 这k个向量将被用作新空间的基

Problem Formulation（P10）

Can simply subtract each x_i by the sample mean.

- Data representations

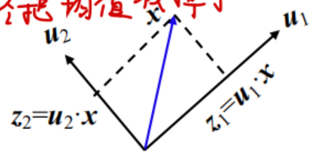
  $X = [x_1, \ldots, x_N],\ x_i \in \mathbb{R}^d$, where $\sum_{i=1}^{N} x_i = 0$



- Projection of a data point

  $z_j = \boldsymbol{u}_j^{\mathrm{T}} x \qquad \boldsymbol{Z} = \boldsymbol{u}^{\mathrm{T}} \boldsymbol{X}$ 投影

已经把均值减掉了

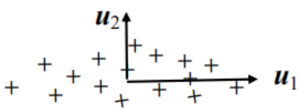


- Variance in the projected space

  方差

  $$var(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{X}) = \tfrac{1}{N}(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{X})(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{X})^{\mathrm{T}}$$
  $$= \tfrac{1}{N}\, \boldsymbol{u}^{\mathrm{T}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\boldsymbol{u}$$
  $$= \boldsymbol{u}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{u}$$

  $S = \mathrm{cov}(X) = \tfrac{1}{N} XX^{\mathrm{T}}$ 原始数据的协方差



- In summary, PCA aims to solve a constrained optimization problem:

$$\max_u \ \boldsymbol{u}^T \underbrace{\boldsymbol{S}}\boldsymbol{u} \quad \text{半正定}$$
$$\text{s.t.} \ \ \boldsymbol{u}^T\boldsymbol{u} = \mathbf{1}$$

- This is a quadratic programming (QP) problem, which is one type of convex optimization problem.
- S is positive semi-definite.

- Lagrangian: $\mathcal{L} = \boldsymbol{u}^T\boldsymbol{S}\boldsymbol{u} - \lambda(\boldsymbol{u}^T\boldsymbol{u} - 1)$

$$\nabla_{\boldsymbol{u}}\mathcal{L} = 0 \ \Rightarrow \ \underline{\boldsymbol{S}\boldsymbol{u} = \lambda\boldsymbol{u}}$$ 特征值，u 为特征向量

- Solving PCA amounts to **eigen decomposition**:

  $\boldsymbol{u}$: eigen vector, λ: eigen value

- Since $\boldsymbol{S}$ is positive semi-definite, we have

$$\boldsymbol{S} = \boldsymbol{U\Sigma U}^T = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_d]\mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_d)[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_d]^T$$

- Find k orthogonal principal components

  correspond to the k eigen vectors with k largest eigen values

$$\mathrm{S} = \boldsymbol{U\Sigma U}^T \approx \boldsymbol{U}_{1:k}\boldsymbol{\Sigma}_{1:k}\boldsymbol{U}^T_{1:k}$$

- Project each input vector x into this subspace

$$z_j = \boldsymbol{u}_j^T x \qquad z = \boldsymbol{U}_{1:k}{}^T\boldsymbol{x}$$

Intuitions:

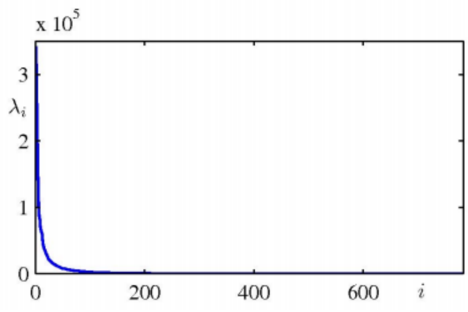
– *Eigen value is the variance of the data projected to the corresponding eigen vector*

– *Eigen vectors are essentially orthogonal* 正交

K的选取规则：信息损失尽可能少（P14）

- **What to discard**: the eigen vectors with small eigen values
- Limit the loss of information within an acceptable interval (usually 5%)

$$\text{info loss} = \frac{\lambda_{k+1} + \ldots + \lambda_d}{\lambda_1 + \lambda_2 + \ldots + \lambda_d}$$



Variance preserved at i-th eigenvalue

Algorithm（P15）

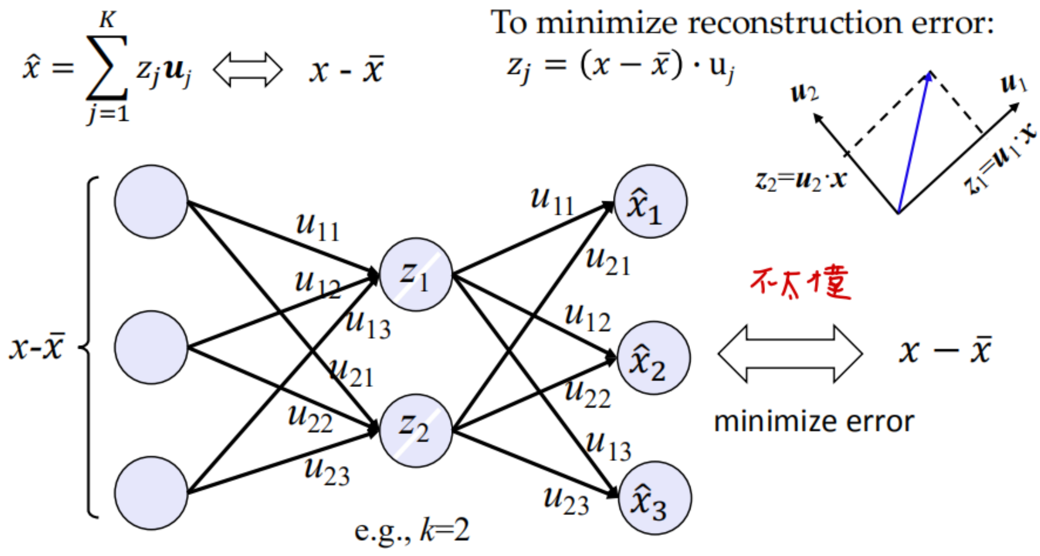
1. Shifting data set to have zeros mean
2. Compute the covariance matrix S
3. Conduct eigen decomposition on S, and rank the eigen vectors according to their eigen values
4. Determine the number of dimensions k of the new space
5. Select the first k eigen vectors with d largest eigen values as the basis of the new space

手工计算示例：(79条消息) 利用PCA降维的手工计算实例_独孤呆博的博客-CSDN博客_pca降维举例

Applications of PCA（P17-19）

- Data visualization
- Preprocessing
- Modeling – prior for new data 先验
- Compression 压缩

PCA和neural network等价

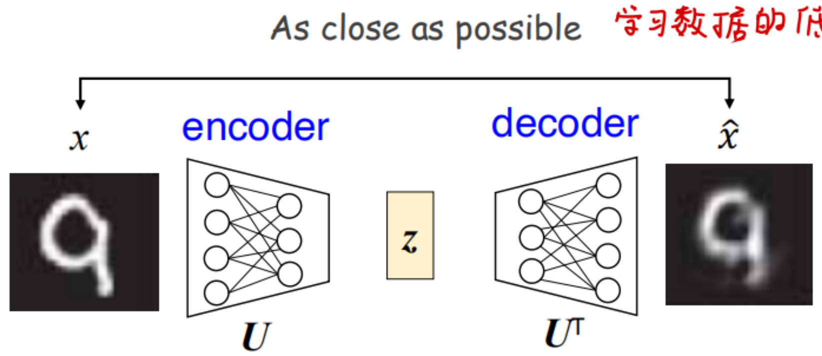


PCA = a neural network with one hidden layer (linear activation)

auto-encoder

PCA等价于一层的auto encoder

- **Auto-encoder**: an encoder-decoder **neural network** whose output tries to reconstruct the input.

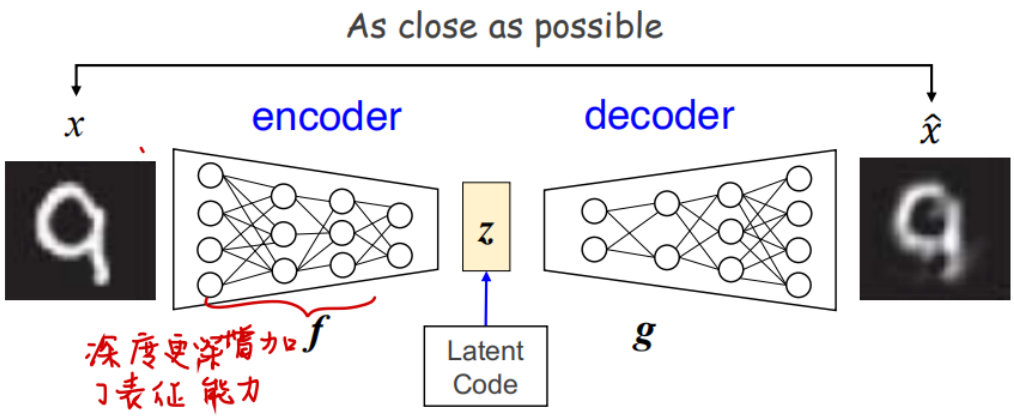


学习数据的低维表征

目标 $\hat{x}$ 和 x 越近越好

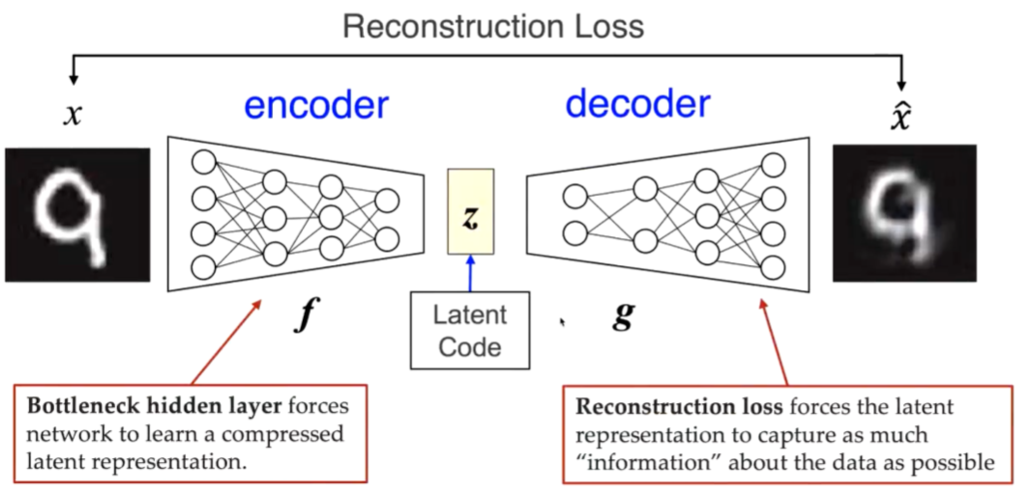
U：主成分
z:对应的投影

- Of course, the auto-encoder can be deep



  - ▻ symmetric is not necessary.
  - ▻ can be initialized layer-wise by RBM 受限玻尔兹曼机

- Of course, the auto-encoder can be deep



- encoder： 瓶颈在于数据的压缩
- decoder: 尽可能保持原来的信息

Denoising Auto-encoder（P26）

用于降噪

- An autoencoder that receives a **corrupted** data point as input and is trained to predict the original data point.