Lecture 14: Feature extraction

onhie Rober

Introduction

Principal Component Analysis

Principle
Algorithm
Advantages an

Advantages and drawbacks

Principle
Example

Lecture 14: Feature extraction Introduction to Machine Learning

Sophie Robert

L3 MIASHS — Semestre 2

2022-2023

Principle Algorithm

Advantages and drawbacks

Principle
Example
Advantages:

1 Introduction

- 2 Principal Component Analysis
 - Principle
 - Algorithm
 - Advantages and drawbacks
- 3 t-SNE
 - Principle
 - Example
 - Advantages and drawbacks

Lecture 14: Feature extraction

ophie Robe

Introduction

Principal Component Analysis

Principle Algorithm

Advantages and drawbacks

Principle Example Advantages an

Question

Do you remember what is the definition of feature extraction/projection ?

Lecture 14: Feature extraction

Introduction

Question

Do you remember what is the definition of feature extraction/projection?

■ Find a lower dimensional space to project the data in

Lecture 14: Feature extraction

ophie Robe

Introduction

Principal Componen Analysis

Principle Algorithm Advantages an drawbacks

drawbacks

Principle
Example
Advantages and

Question

Do you remember what is the definition of feature extraction/projection ?

- Find a lower dimensional space to project the data in
- While keeping as much information as possible

Lecture 14: Feature extraction

Introduction

Many methods exist in the literature:

Lecture 14: Feature extraction

ophie Robei

Introduction

Principal Component Analysis

Principle
Algorithm

Advantages ai drawbacks

Principle
Example

Many methods exist in the literature:

■ Linear transformations: PCA, . . .

Lecture 14: Feature extraction

ophie Robe

Introduction

Component Analysis Principle Algorithm Advantages and drawbacks

t-SNE
Principle
Example
Advantages and

Many methods exist in the literature:

- Linear transformations: PCA, . . .
- Non-linear transformations: t-distributed stochastic neighbourhood embedding (t-SNE), Uniform Manifold Approximation and Projection (UMAP) ...

Lecture 14: Feature extraction

Sopille Rober

Principal

Component Analysis

Principle
Algorithm
Advantages a

Advantages and drawbacks

t-SNE

Example
Advantages an

Principal Component Analysis

Principal Component analysis is a feature projection method that consists in finding a new coordinate system as a linear combination of the input features to project the data: this system is orthogonal and a linear combination of the features that maximizes the variance.

Lecture 14: Feature extraction

Principle

Principal Component Analysis

Principal Component analysis is a feature projection method that consists in finding a new coordinate system as a linear **combination of the input features** to project the data: this system is orthogonal and a linear combination of the features that maximizes the variance.

The goal of PCA is to:

Lecture 14: Feature extraction

Principal Componer

Principle
Algorithm
Advantages an

Advantages an drawbacks

Principle

Example
Advantages and drawbacks

Principal Component Analysis

Principal Component analysis is a feature projection method that consists in finding a new coordinate system as a linear combination of the input features to project the data: this system is orthogonal and a linear combination of the features that maximizes the variance.

The goal of PCA is to:

Find new axis more relevant to represent the data

Lecture 14: Feature extraction

Principal Componer Analysis

Principle
Algorithm
Advantages as

Advantages and drawbacks

Principle
Example
Advantages a

Principal Component Analysis

Principal Component analysis is a feature projection method that consists in finding a new coordinate system as a linear combination of the input features to project the data: this system is orthogonal and a linear combination of the features that maximizes the variance.

The goal of PCA is to:

- Find new axis more relevant to represent the data
- Give us the importance of each axis

Lecture 14: Feature extraction

Principal Componer Analysis

Principle
Algorithm
Advantages andrawbacks

drawbacks

Principle Example Advantages and drawbacks

Principal Component Analysis

Principal Component analysis is a feature projection method that consists in finding a new coordinate system as a linear combination of the input features to project the data: this system is orthogonal and a linear combination of the features that maximizes the variance.

The goal of PCA is to:

- Find new axis more relevant to represent the data
- Give us the importance of each axis
- Remove unimportant axis and reduce dimension

Lecture 14: Feature extraction

Introduction

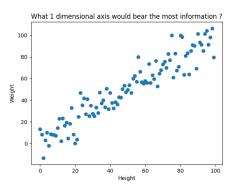
Principal Componen

Principle

Advantages an

drawbacks

Principle
Example
Advantages and



Question

What would be an adequate axis to project the data on to reduce to a single axis?

Lecture 14: Feature extraction

Sophie Robert

Introductio

Componer Analysis

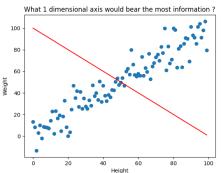
Principle

Advantages an

drawbacks

Principle
Example
Advantages

Possibility: y = -x



Question

What would a projection on this new axis look like?

Lecture 14: Feature extraction

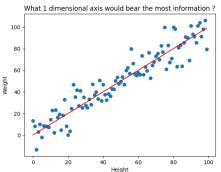
Principal

Componer Analysis

Principle

Advantages and drawbacks

Principle Example Possibility: y = x



Question

What would a projection on this new axis look like?

Lecture 14: Feature extraction

Principle

We are looking for new axis that maximize the variance and are uncorrelated.

Lecture 14: Feature extraction

_ . . .

Componen Analysis Principle

Algorithm
Advantages a

drawbacks

Principle Example Advantages an We are looking for new axis that **maximize the variance** and are **uncorrelated**.

It can be shown that these **principal components** are eigenvectors of the data's:

Lecture 14: Feature extraction

phie Robert

Introduction

Principal Componen Analysis Principle

Principle Algorithm Advantages an drawbacks

t-SNE
Principle

We are looking for new axis that **maximize the variance** and are **uncorrelated**.

It can be shown that these **principal components** are eigenvectors of the data's:

Covariance matrix (Centered PCA)

Lecture 14: Feature extraction

phie Robert

Introduction

Principal Componen Analysis

Principle
Algorithm
Advantages ar

Advantages an drawbacks t-SNE

Principle
Example
Advantages an

We are looking for new axis that **maximize the variance** and are **uncorrelated**.

It can be shown that these **principal components** are eigenvectors of the data's:

- Covariance matrix (Centered PCA)
- Correlation matrix (Normed PCA)

Example dataset

Lecture 14: Feature extraction

Sopnie Robert

Principal Componen

Analysis

Principle Algorithm

Advantages a

drawbacks

Principle

Example Advantages and drawbacks

The Iris dataset <i>D</i> :		
sepal length (cm)	sepal width (cm)	petal length (cm)
5.1	3.5	1.4
4.9	3.0	1.4
4.7	3.2	1.3
4.6	3.1	1.5
5.0	3.6	1.4
5.4	3.9	1.7
4.6	3.4	1.4
5.0	3.4	1.5
4.4	2.9	1.4
4.9	3.1	1.5

Example dataset

Lecture 14: Feature extraction

Sophie Robe

Principal Componen

Principle

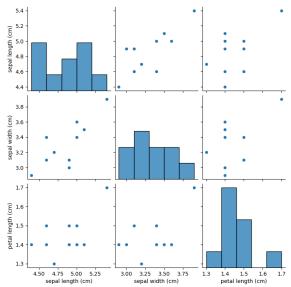
Algorithm

Advantages ar drawbacks

+ CNE

Principle Example

Example Advantages and drawbacks



Standardize matrix

Lecture 14: Feature extraction

Algorithm

For **centered PCA**, center matrix. For normed PCA, standardize matrix (remove mean and divide by standard error): Z.

Compute the correlation matrix

Lecture 14: Feature extraction

Compute correlation matrix C.

Principal Componer

Principle

Algorithm

Advantages and

drawbacks t-SNE

Principle
Example
Advantages and drawbacks

With Z the standardized matrix and n the number of individuals within the dataset, $C = \frac{1}{n}Z^tZ$

	Sepal length	width	Petal length
Sepal length	1.00	0.79	0.60
width	0.79	1.00	0.52
Petal length	0.60	0.52	1.00

Find eigenvalues and eigenvectors

Lecture 14: Feature extraction

hie Robert

Principal Componen

Principle
Algorithm

Advantages and drawbacks

t-SNE Principle Example Advantages and Eigen values (sorted) of the correlation matrix are: [2.27, 0.51, 0.20]

Eigen vectors matrix P (sorted by eigen values) is:

Feature	Component 0	Component 1	Component 2
S. length	0.61	-0.26	0.75
Width	0.59	-0.48	-0.65
P. length	0.53	0.84	-0.14

Project matrix into new space

Lecture 14: Feature extraction

Sophie Robert

Introduction

Principal Component Analysis

Algorithm Advantages an

drawbacks t-SNE

Principle
Example
Advantages

Multiply the standardized matrix Z by the eigenvectors to have the matrix Z^* in the new projected space $Z^* = ZP$: this is the projection of the individuals in the new feature space.

ID	Component 0	Component 1	Component 2
0	0.66	-0.95	0.30
1	-0.80	0.07	0.87
2	-1.35	-0.90	0.02
3	-0.74	1.00	-0.31
4	0.64	-1.02	-0.20
5	3.68	0.57	-0.20
6	-0.65	-0.31	-0.83
7	0.75	0.13	0.11
8	-2.11	0.70	-0.26
9	-0.08	0.72	0.51

Select number of axis

Lecture 14: Feature extraction

ophie Robe

Introduction

Principal Component Analysis

> Principle Algorithm

Advantages an

drawbacks t-SNE

Principle
Example
Advantages and

Axis importance

The value of each eigenvalue is the **importance of the axis** and is used to select the number of axis to keep as a percentage of the explained variance.

Select number of axis

Lecture 14: Feature extraction

Introduction

Principal Componen Analysis

> Principle Algorithm

Advantages and drawbacks

Principle
Example
Advantages

Axis importance

The value of each eigenvalue is the **importance of the axis** and is used to select the number of axis to keep as a percentage of the explained variance.

Usually, compute the cumulated sum of $\frac{\lambda_i}{\sum_{i=0}^n \lambda_i}$ for the i-th eigenvalue λ_i : the selected value is the **total explained** variance of the dataset.

Select number of axis

Lecture 14: Feature extraction

Here the cumulated sum is: [0.76, 0.93, 1]

Introduction

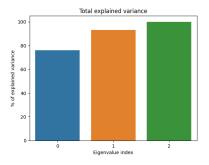
Principal Componen Analysis

Principle Algorithm

Advantages an

t-SNE

Example
Advantages and
drawbacks



We select two axis.

Final projection

Lecture 14: Feature extraction

lukus dirakisis

Principal Componer Analysis

Principle
Algorithm

Advantages an drawbacks

t-SNE

Principle
Example
Advantage

We have our final projection for each individual within the reduced space:

ID	Component 0	Component 1
0	0.66	-0.95
1	-0.80	0.07
2	-1.35	-0.90
3	-0.74	1.00
4	0.64	-1.02
5	3.68	0.57
6	-0.65	-0.31
7	0.75	0.13
8	-2.11	0.70
9	-0.08	0.72

Interpret results

Lecture 14: Feature extraction

A possible interpretation is the plot of the **correlation circle**.

Algorithm

Interpret results

Lecture 14: Feature extraction

Principal Componer

Principle
Algorithm

Advantages and drawbacks

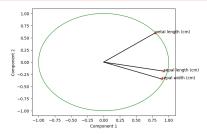
t-SNE

Principle
Example
Advantages and drawbacks

A possible interpretation is the plot of the **correlation circle**.

Correlation circle

The correlation circle consists in computing the correlation of the original features with the new components and deducing from it the contribution of each variable to the axis.



Lecture 14: Feature extraction

hie Robert

Principal

Componer Analysis

Principle Algorithm

Advantages and

drawbacks

urawbacks

Principle

. Advantages and Advantages

Lecture 14: Feature extraction

Advantages and drawbacks

Advantages

Simple to implement

Lecture 14: Feature extraction

ophie Robert

Principal Compone

Principle

Advantages and

drawbacks

Principle Example

Advantages

- Simple to implement
- Simple to evaluate data loss because of transformation

Lecture 14: Feature extraction

ophie Robert

Principal

Componen Analysis

Principle Algorithm

Advantages and drawbacks

drawbacks + SNF

Principle Example Advantages and

Advantages

- Simple to implement
- Simple to evaluate data loss because of transformation

Drawbacks

 Features are transformed and lose interpretability of some results

Lecture 14: Feature extraction

opnie Rober

Principal Componer Analysis

Principle Algorithm

Advantages and drawbacks

drawbacks

Principle
Example
Advantages and

Advantages

- Simple to implement
- Simple to evaluate data loss because of transformation

- Features are transformed and lose interpretability of some results
- Sensitive to outliers

t-distributed Stochastic Neighbor Embedding

Lecture 14: Feature extraction

ophie Robert

Principal Componer Analysis

Principle Algorithm Advantages and drawbacks

Principle
Example
Advantages a

t-SNE

t-distributed Stochastic neighbor embedding (t-SNE) is a nonlinear feature reduction method that consists in associating for each pair of individual **the probability of being close** to each other in the new space by relying on a probability distribution in the original space.

t-distributed Stochastic Neighbor Embedding

Lecture 14: Feature extraction

sopnie Rober

Component Analysis Principle Algorithm Advantages and drawbacks

Principle
Example
Advantages

t-SNE

t-distributed Stochastic neighbor embedding (t-SNE) is a nonlinear feature reduction method that consists in associating for each pair of individual **the probability of being close** to each other in the new space by relying on a probability distribution in the original space.

Given a matrix X with k individuals and n features, find a matrix Q with k individuals and d features, with d << n

Lecture 14: Feature extraction

opine Robei

Introduction

Principal Compone

Analysis
Principle

Algorithm

Advantages : drawbacks

t-SNF

Principle

Advantages and drawbacks We need two probability distributions:

Lecture 14: Feature extraction

Introduction

Principal Componen Analysis

> Principle Algorithm

Advantages ar drawbacks

Principle

Example Advantages and drawbacks We need two probability distributions:

• p_{ij} : measures the similarity between the individuals within the initial space.

Lecture 14: Feature extraction

Principle

We need two probability distributions:

- p_{ii} : measures the similarity between the individuals within the initial space.
- q_{ii} : measures the similarity between the individuals within the new space.

Lecture 14: Feature extraction

oop...c ..obc.

Principal

Componen Analysis Principle

Algorithm Advantages ar drawbacks

t-SNE Principle

Principle Example Advantages and drawbacks We need two probability distributions:

- p_{ij} : measures the similarity between the individuals within the initial space.
- q_{ij} : measures the similarity between the individuals within the new space.

The divergence between these probability distributions **needs** to **minimized**.

Kullback-Leibler divergence

Lecture 14: Feature extraction

Introduction

Principal Componen Analysis

Principle Algorithm Advantages an drawbacks

drawbacks

Principle Example Advantages

Kullback-Leibler divergence

The **Kullback-Leibler divergence** is a statistical similarity that measures how one probability distribution P is different from a distribution Q.

$$D_{KL}(P||Q) = \sum_{x \in \mathcal{X}} P(x) log(\frac{P(x)}{Q(x)})$$

In the case of t-SNE we want to minimize:

$$\mathit{KL}(P||Q) = \sum_{i \neq j} p_{ij} log(\frac{p_{ij}}{q_{ij}})$$

Lecture 14: Feature extraction

Principle

We define the conditional probability of j being the neighbor of i as a Gaussian around x_i (σ_i measures the density around x_i and should be estimated):

$$p_{j|i} = \frac{\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|\mathbf{x}_i - \mathbf{x}_k\|^2 / 2\sigma_i^2)}$$

Lecture 14: Feature extraction

Principle

We define the conditional probability of *i* being the neighbor of i as a Gaussian around x_i (σ_i measures the density around x_i and should be estimated):

$$p_{j|i} = \frac{\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|\mathbf{x}_i - \mathbf{x}_k\|^2 / 2\sigma_i^2)}$$

and set:

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N}$$

Lecture 14: Feature extraction

Introduction

Principal Componen

Principle Algorithm Advantages and drawbacks

t-SNE Principle Example

Principle Example Advantages and drawbacks We define the conditional probability of j being the neighbor of i as a Gaussian around x_i (σ_i measures the density around x_i and should be estimated):

$$\rho_{j|i} = \frac{\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|\mathbf{x}_i - \mathbf{x}_k\|^2 / 2\sigma_i^2)}$$

and set:

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N}$$

In the new dimensional space, we set a heavy-tailed Student law:

Lecture 14: Feature extraction

Introduction

Principal Componen

Analysis
Principle
Algorithm

Advantages and drawbacks

Principle
Example
Advantages

We define the conditional probability of j being the neighbor of i as a Gaussian around x_i (σ_i measures the density around x_i and should be estimated):

$$\rho_{j|i} = \frac{\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|\mathbf{x}_i - \mathbf{x}_k\|^2 / 2\sigma_i^2)}$$

and set:

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N}$$

In the new dimensional space, we set a heavy-tailed Student law:

$$q_{ij} = \frac{(1 + \|\mathbf{q}_i - \mathbf{q}_j\|^2)^{-1}}{\sum_k \sum_{l \neq k} (1 + \|\mathbf{q}_k - \mathbf{q}_l\|^2)^{-1}}$$

Lecture 14: Feature extraction

Introduction

Principal Componen

Principle
Algorithm
Advantages and

t-SNE Principle

Principle Example Advantages and drawbacks We define the conditional probability of j being the neighbor of i as a Gaussian around x_i (σ_i measures the density around x_i and should be estimated):

$$p_{j|i} = \frac{\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|\mathbf{x}_i - \mathbf{x}_k\|^2 / 2\sigma_i^2)}$$

and set:

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N}$$

In the new dimensional space, we set a heavy-tailed Student law:

$$q_{ij} = \frac{(1 + \|\mathbf{q}_i - \mathbf{q}_j\|^2)^{-1}}{\sum_k \sum_{l \neq k} (1 + \|\mathbf{q}_k - \mathbf{q}_l\|^2)^{-1}}$$

and use gradient descent to minimize the KL divergence.

Example: Iris dataset

Lecture 14: Feature extraction

Sophie Robert

luano di casta a

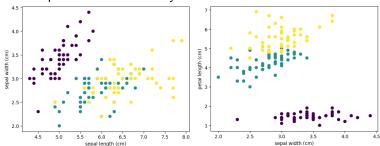
Principal Componen

Principle Algorithm Advantages a

Advantages ar drawbacks

Principle Example

Data repartition without any transformation:



Example: Iris dataset

Lecture 14: Feature extraction

Introduction

Principal Componen Analysis

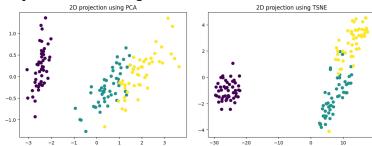
Principle
Algorithm
Advantages

Advantages ar drawbacks

Principle

Example
Advantages and

Projection in 2-D using t-SNE and PCA:



Lecture 14: Feature extraction

Advantages and drawbacks

Advantages:

Lecture 14: Feature extraction

Advantages and drawbacks

Advantages:

Usually visually nicer than PCA

Lecture 14: Feature extraction

Advantages and drawbacks

Advantages:

- Usually visually nicer than PCA
- Can find pattern in data that is non-linear

Lecture 14: Feature extraction

sopille Roberi

Principal

Componen Analysis

Algorithm

Advantages as

Advantages ar drawbacks

Principle Example

Advantages and drawbacks

Advantages:

- Usually visually nicer than PCA
- Can find pattern in data that is non-linear

Lecture 14: Feature extraction

Advantages and drawbacks

Advantages:

- Usually visually nicer than PCA
- Can find pattern in data that is non-linear

Drawbacks:

Many hyperparameters

Lecture 14: Feature extraction

Advantages and drawbacks

Advantages:

- Usually visually nicer than PCA
- Can find pattern in data that is non-linear

- Many hyperparameters
- Only preserves local structure of the data

Lecture 14: Feature extraction

Advantages and drawbacks

Advantages:

- Usually visually nicer than PCA
- Can find pattern in data that is non-linear

- Many hyperparameters
- Only preserves local structure of the data
- Stochastic results

Lecture 14: Feature extraction

Principal

Analysis Principle

Algorithm Advantages an drawbacks

drawbacks

Principle
Example
Advantages and

Advantages:

- Usually visually nicer than PCA
- Can find pattern in data that is non-linear

- Many hyperparameters
- Only preserves local structure of the data
- Stochastic results
- Does not scale well

Lecture 14: Feature extraction

Jopine Robert

Principal

Componen Analysis

Principle Algorithm Advantages

Advantages and drawbacks

Principle Example

Example
Advantages and
drawbacks

Advantages:

- Usually visually nicer than PCA
- Can find pattern in data that is non-linear

- Many hyperparameters
- Only preserves local structure of the data
- Stochastic results
- Does not scale well
- Hard to interpret

Lecture 14: Feature extraction

Principal Componer

Analysis Principle Algorithm

Advantages and drawbacks

t-SNE Principle Example

Example Advantages and drawbacks

Advantages:

- Usually visually nicer than PCA
- Can find pattern in data that is non-linear

- Many hyperparameters
- Only preserves local structure of the data
- Stochastic results
- Does not scale well
- Hard to interpret
- Does not create new variables that can be understood

Questions

Lecture 14: Feature extraction

nie Robert

Introductio

Principal Componer

Analysis Principle

Algorithm

Advantages au

drawbacks

Principle Example

Advantages and drawbacks

Questions ?