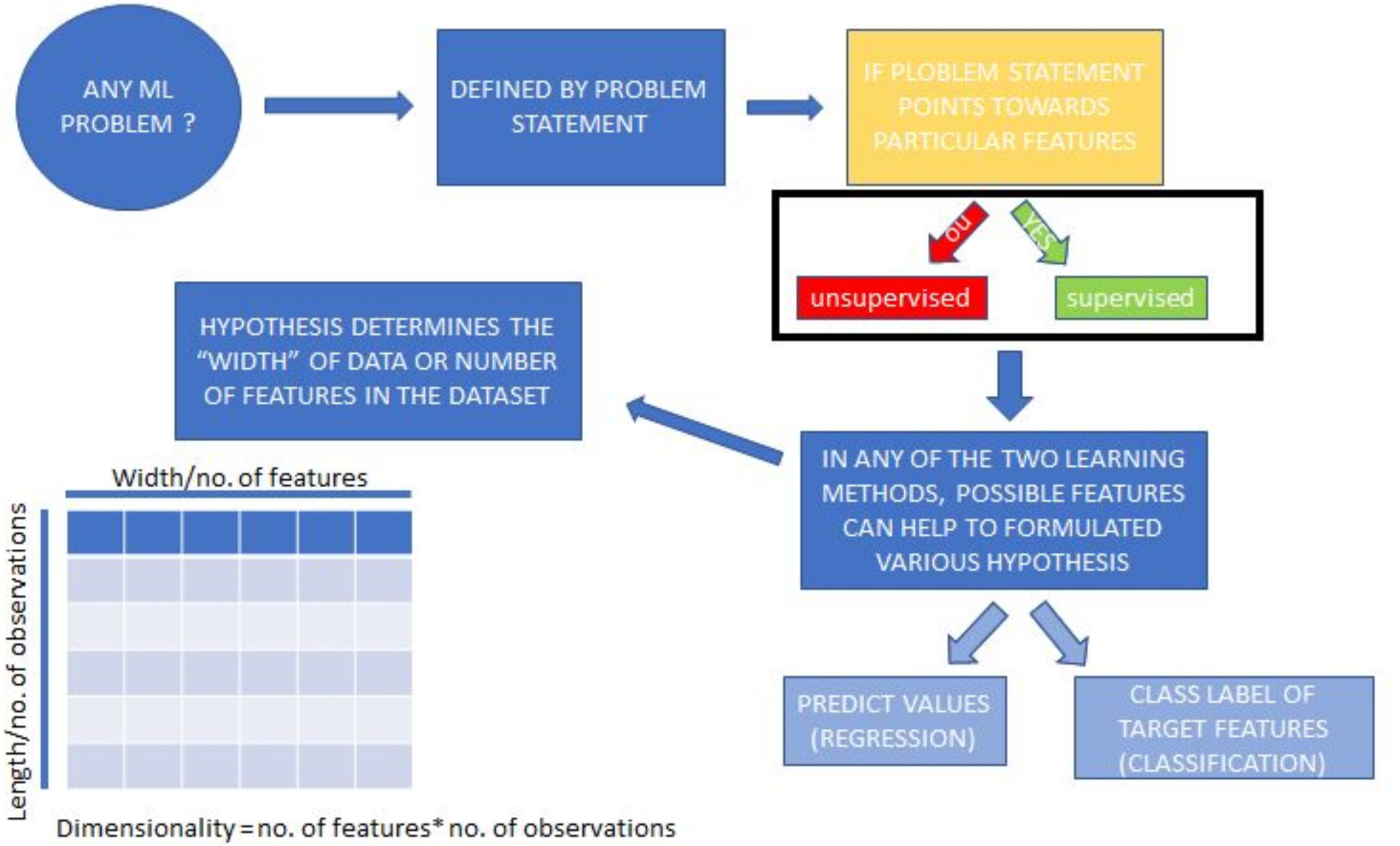
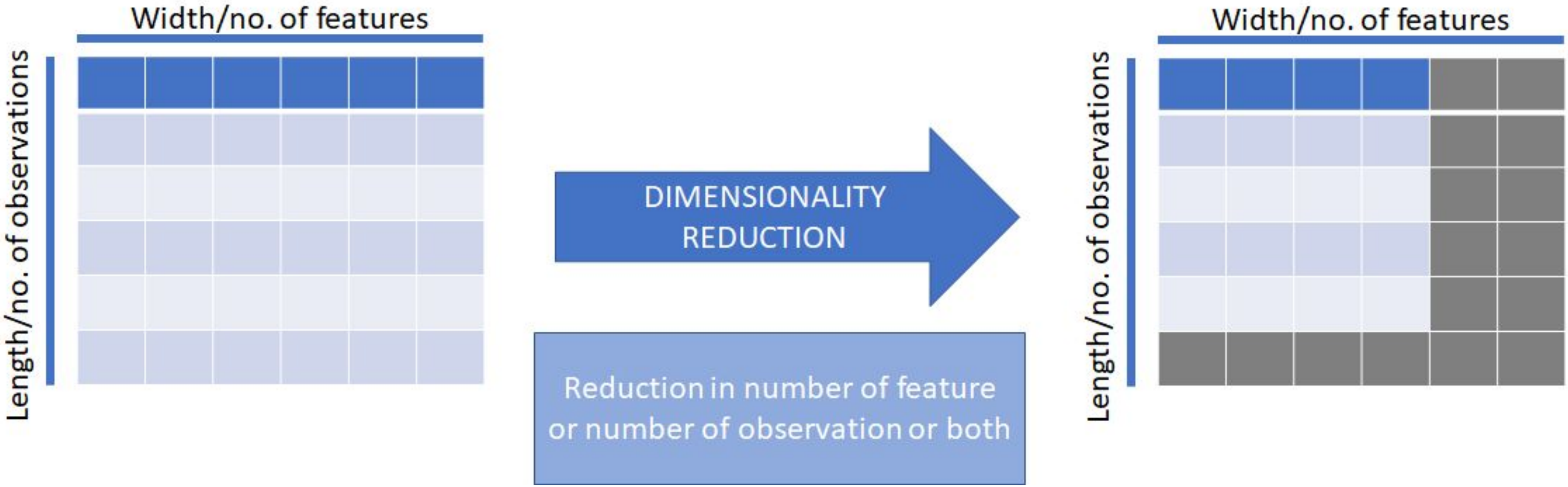
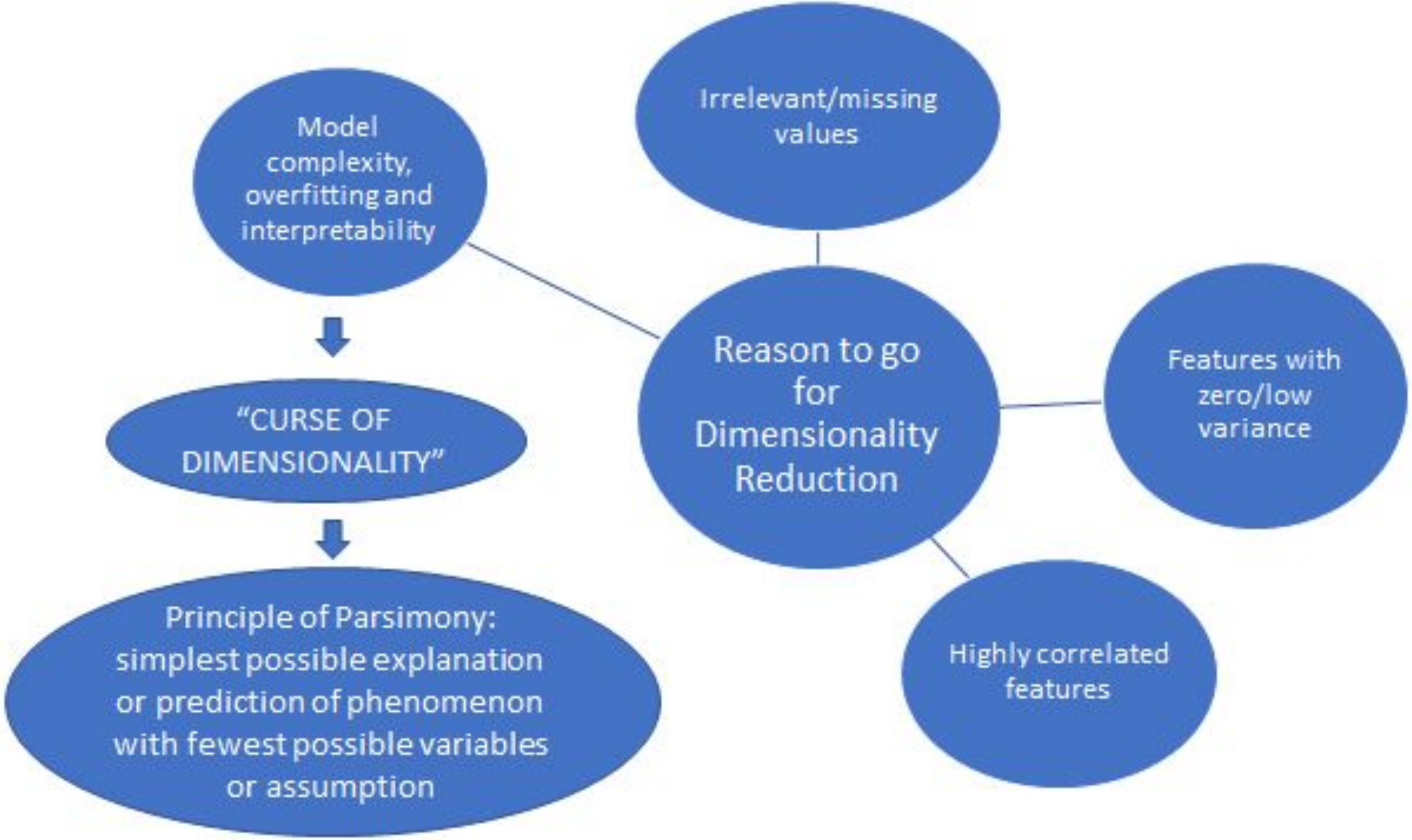
RECAP CONCEPT OF DIMENSIONALITY REDUCTION



RECAP CONCEPT OF DIMENSIONALITY REDUCTION




Model complexity, overfitting and interpretability
"CURSE OF DIMENSIONALITY"
Principle of Parsimony: simplest possible explanation or prediction of phenomenon with fewest possible variables or assumption
Irrelevant/missing values
Reason to go for Dimensionality Reduction
Features with zero/low variance
Highly correlated features

Approach for Dimensionality Reduction

- Pre-processing
  - Feature Selection or Elimination/Data Compression: Data preprocessing stage or prior to predictive modelling stage
  - Feature Importance/Noise Reduction: objective is to select the top independent features that contribute to variations in the target features
- Dimensionality Reduction
  - Data Classification
  - Data Visualization: Includes Projection and Manifold learning (example PCA, LDA, t-SNE and UMAP)

1. Preprocessing
2. Dimensionality reduction

Preprocessing

- dimensions of the data
- type of data
- outliers
- missing values

Preprocessing

- **dimensions of the data**

  feature selection / dimensionality reduction?

- **type of data**

- **outliers**

- **missing values**

Preprocessing

- **dimensions of the data**

  feature selection / dimensionality reduction?

- **type of data**

  categorical, numerical (transforming? binning?)

  balanced classes? normally distributed?

- **outliers**

- **missing values**

Preprocessing

- **dimensions of the data**

    feature selection / dimensionality reduction?

- **type of data**

    categorical, numerical (transforming? binning?)

    balanced classes? normally distributed?

- **outliers**

- **missing values**

Preprocessing

- **dimensions of the data**

    feature selection / dimensionality reduction?

- **type of data**

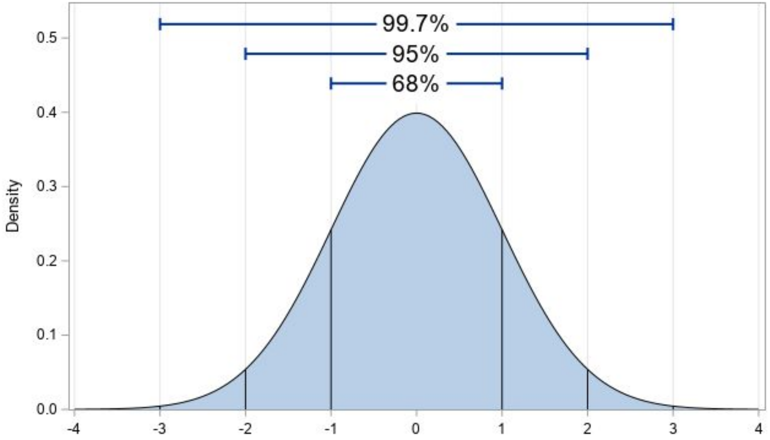
    categorical, numerical (transforming? binning?)

    balanced classes? normally distributed?

- **outliers**

- **missing values**

Density plot



Preprocessing

- **dimensions of the data**

  feature selection / dimensionality reduction?

- **type of data**

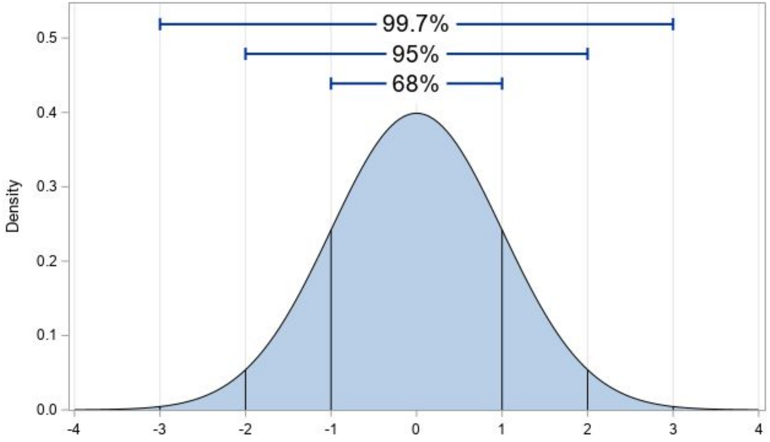
  categorical, numerical (transforming? binning?)

  balanced classes? normally distributed?

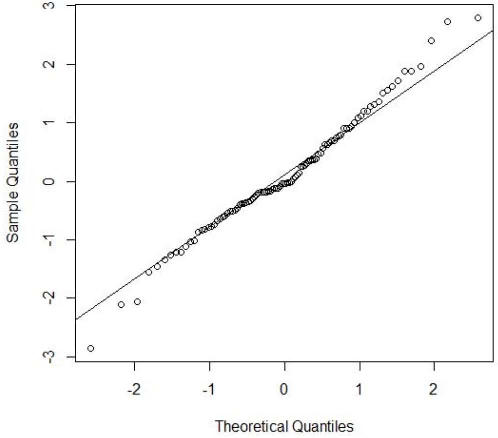
- **outliers**

- **missing values**

Density plot



Q-Q plot



Preprocessing

- **dimensions of the data**

    feature selection / dimensionality reduction?

- **type of data**

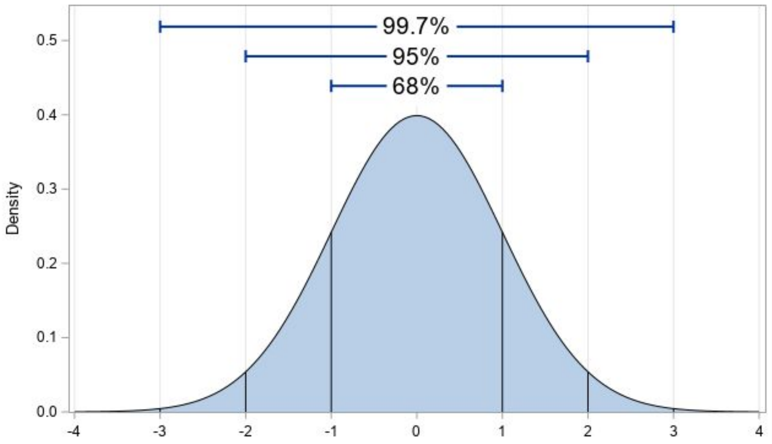
    categorical, numerical (transforming? binning?)

    balanced classes? normally distributed?

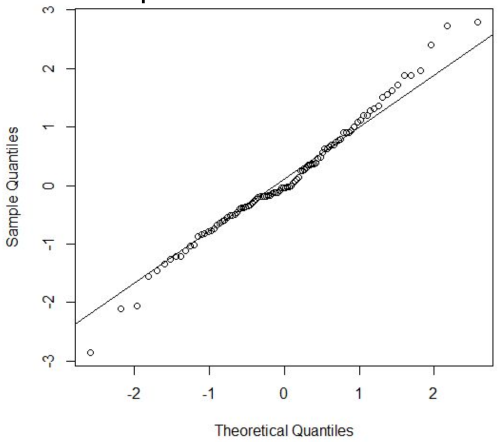
- **outliers**

- **missing values**

Density plot



Q-Q plot



statistical tests (Shapiro–Wilk)

Preprocessing

- **dimensions of the data**

  feature selection / dimensionality reduction?

- **type of data**

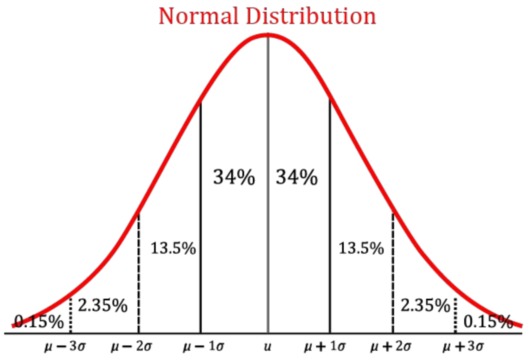
  categorical, numerical (transforming? binning?)

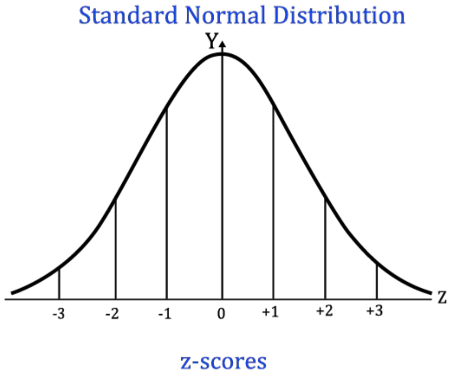
  balanced classes? normally distributed?

- **outliers**

  z-score $$z = \frac{x - \mu}{\sigma}$$

- **missing values**





Preprocessing

- **dimensions of the data**

  feature selection / dimensionality reduction?

- **type of data**

  categorical, numerical (transforming? binning?)

  balanced classes? normally distributed?

- **outliers**

  z-score

- **missing values**

  imputation? remove?

Preprocessing

- **dimensions of the data**

  feature selection / dimensionality reduction?

- **type of data**

  categorical, numerical (transforming? binning?)

  balanced classes? normally distributed?

- **outliers**

  z-score

- **missing values**

  imputation? remove?

HANDS-ON

1. Preprocessing
2. Dimensionality reduction

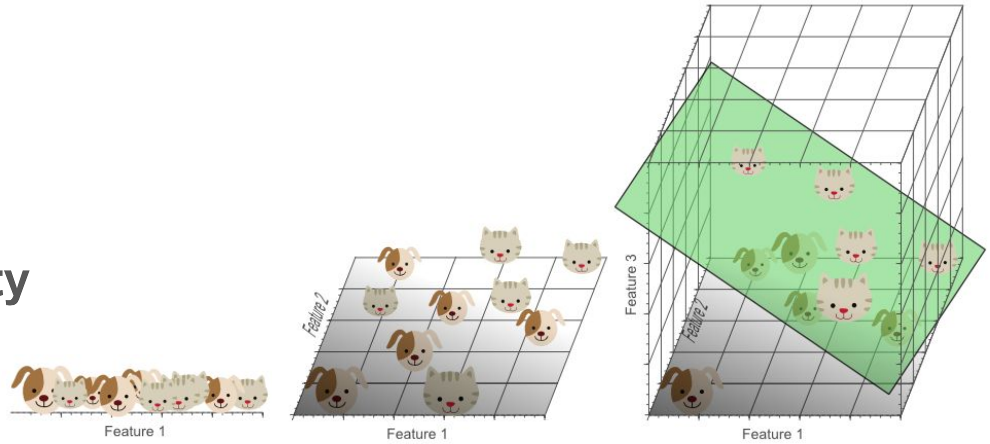
Dimensionality reduction

- **performance vs. interpretability**

    exploratory analysis

- **the curse of dimensionality**

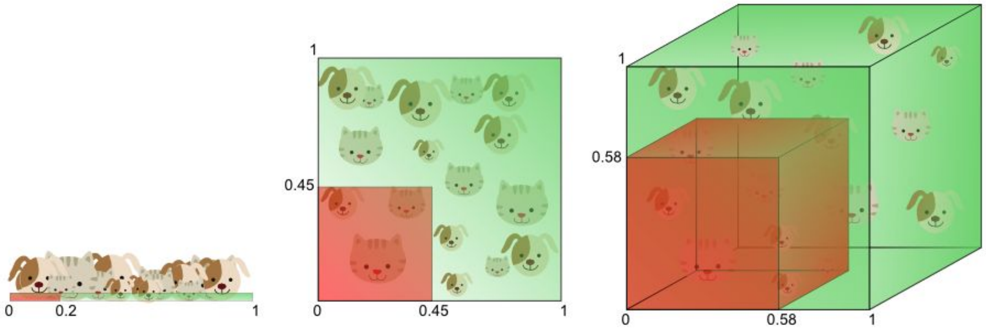
- **overfitting**



Credit: [Vincent Spruyt]

Dimensionality reduction

- **performance vs. interpretability**
    - exploratory analysis
- **the curse of dimensionality**
    - sparseness
- **overfitting**



Credit: Vincent Spruyt

Dimensionality reduction

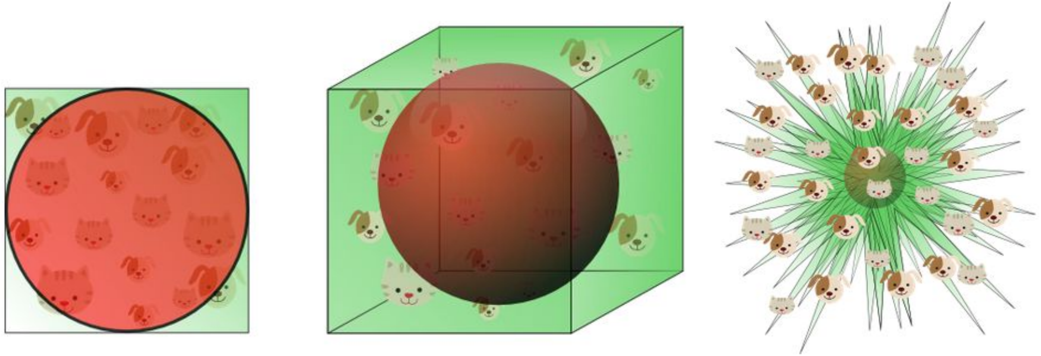
- **performance vs. interpretability**

    exploratory analysis

- **the curse of dimensionality**

    non-uniform sparseness

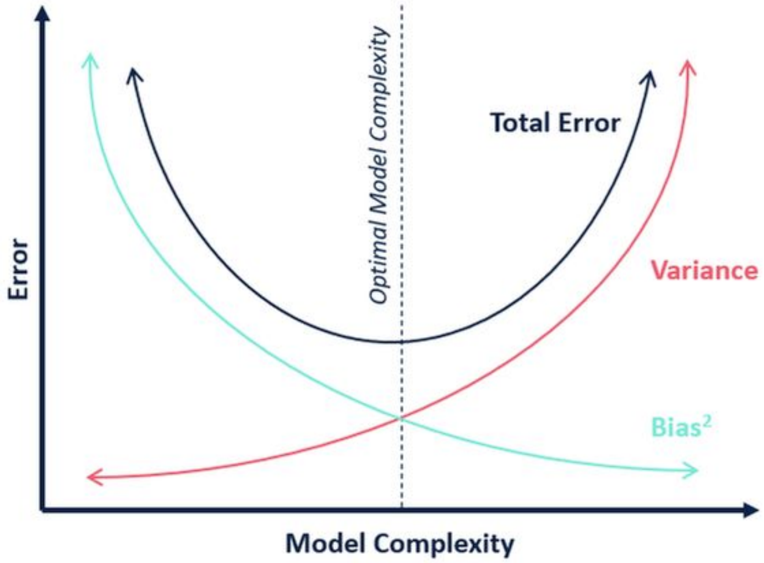
- **overfitting**

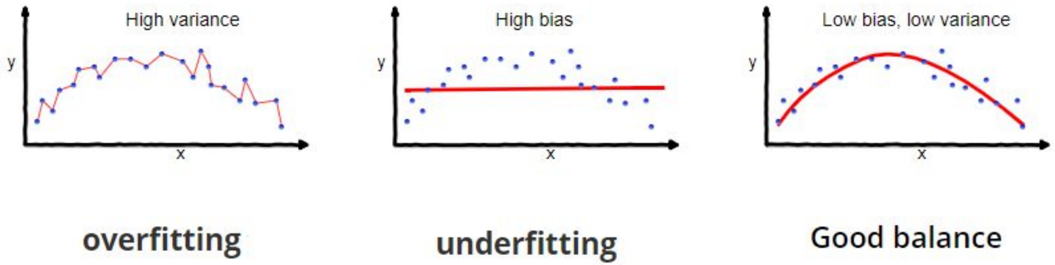


Credit: [Vincent Spruyt]

Dimensionality reduction

- **performance vs. interpretability**
  - exploratory analysis
- **the curse of dimensionality**
  - sparseness
- **overfitting**
  - bias vs variance tradeoff





Dimensionality reduction

- **performance vs. interpretability**

  exploratory analysis

- **the curse of dimensionality**

  sparseness

- **overfitting**

  bias vs variance error

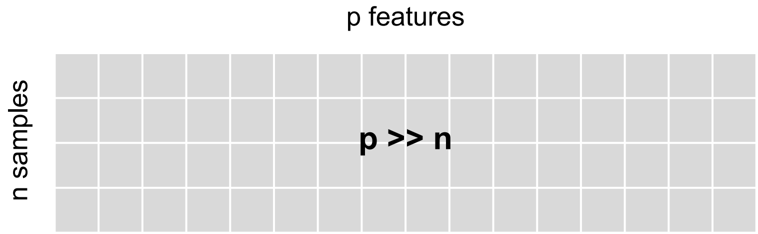
Dimensionality reduction

- **Feature selection**

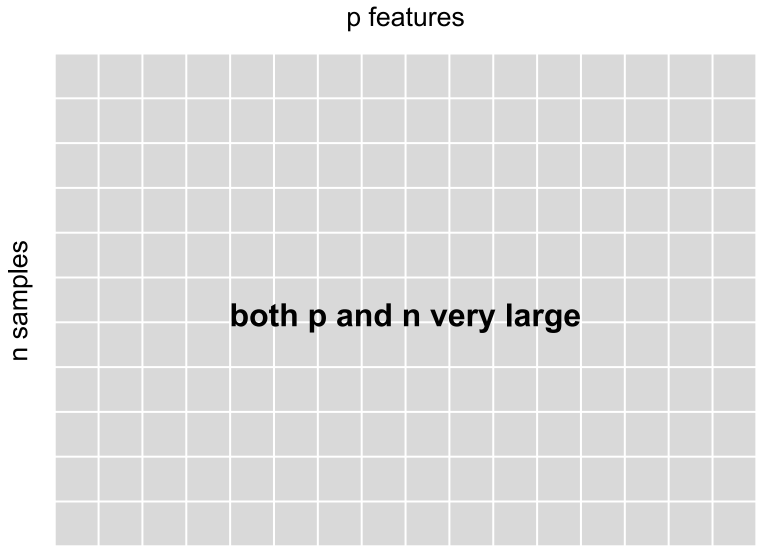
- **Feature extraction**

Dimensionality reduction

- **Feature selection**

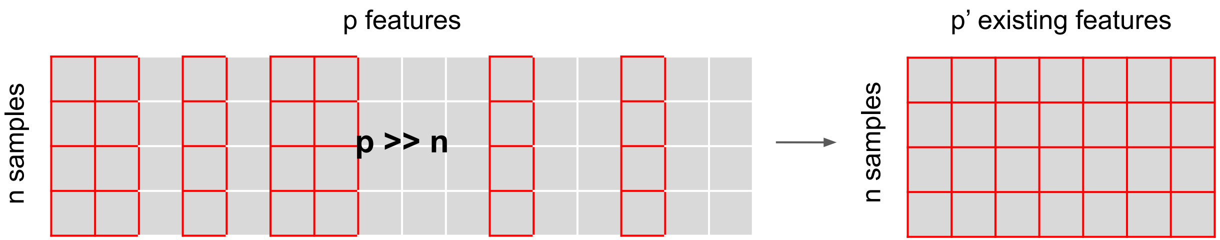


- **Feature extraction**

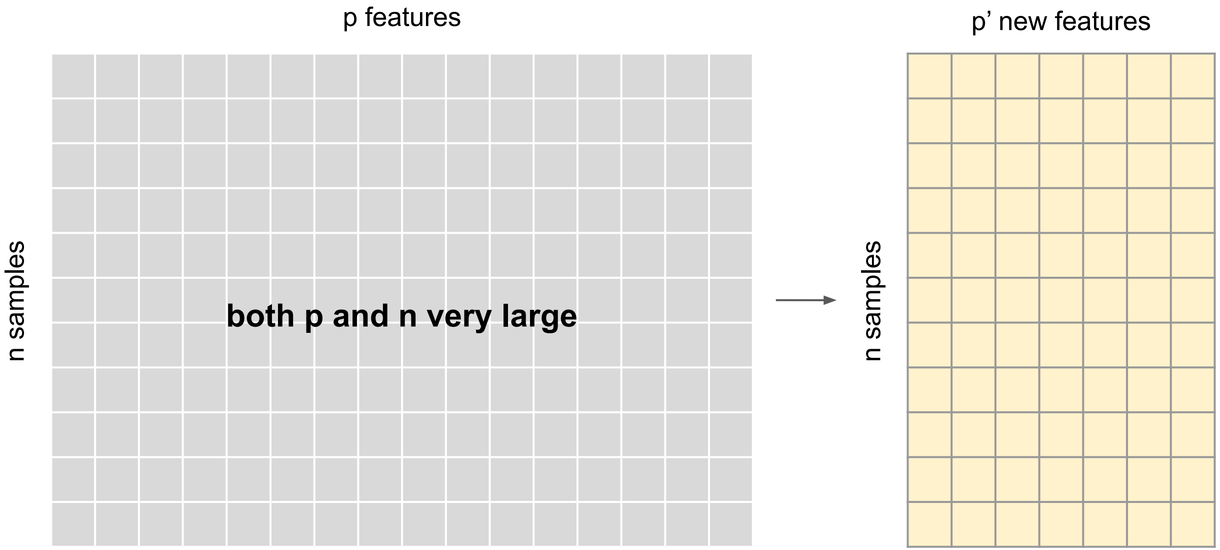


Dimensionality reduction

- **Feature selection**



- **Feature extraction**



Dimensionality reduction

- **Feature selection**

- **Feature extraction**

Dimensionality reduction

- **Feature selection**

    variance feature selection

    regularization: lasso, ridge, elastic net

- **Feature extraction**

Dimensionality reduction

- **Feature selection**

  variance feature selection

  regularization: lasso, ridge, elastic net

- **Feature extraction**

$$RSS = \sum_{i=1}^{N} (\hat{y} - y)^2$$

$$\hat{y} = \beta_0 + \sum_{i=1}^{N} \beta_i x$$

LASSO (L1):

$$RSS + \lambda \sum_{i=1}^{N} |\beta_i|$$

RIDGE (L2):

$$RSS + \lambda \sum_{i=1}^{N} \beta_i^2$$

Dimensionality reduction

- **Feature selection**

  variance feature selection

  regularization: lasso, ridge, elastic net

- **Feature extraction**

HANDS-ON

$$RSS = \sum_{i=1}^{N} (\hat{y} - y)^2$$

$$\hat{y} = \beta_0 + \sum_{i=1}^{N} \beta_i x$$

LASSO (L1):

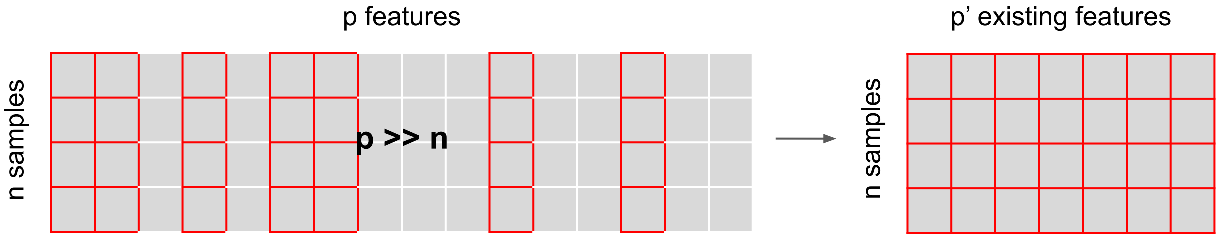
$$RSS + \lambda \sum_{i=1}^{N} |\beta_i|$$

RIDGE (L2):

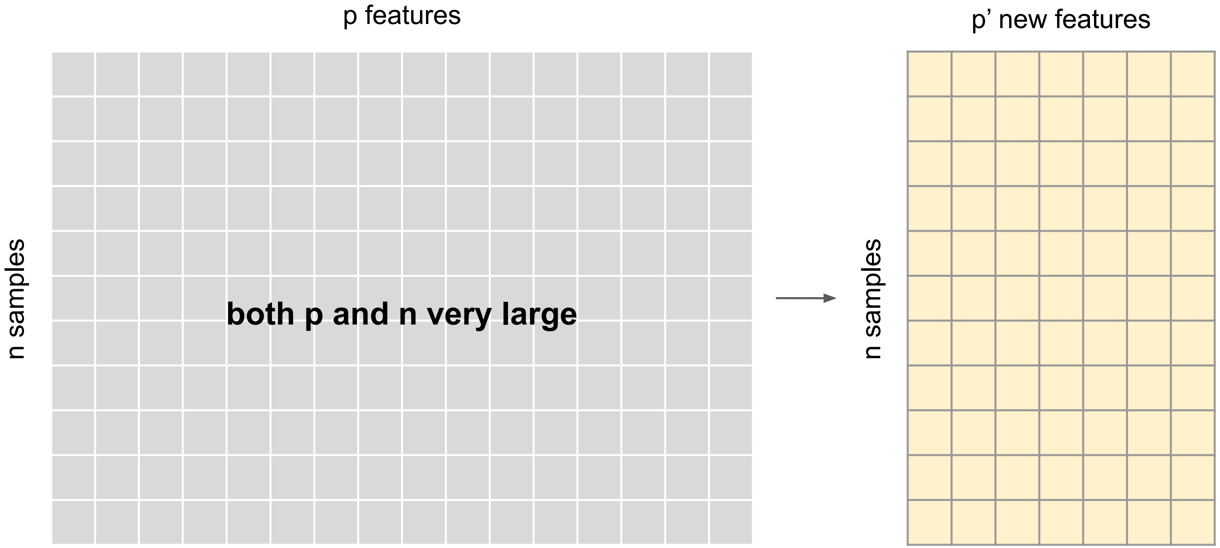
$$RSS + \lambda \sum_{i=1}^{N} \beta_i^2$$

Dimensionality reduction

- **Feature selection**

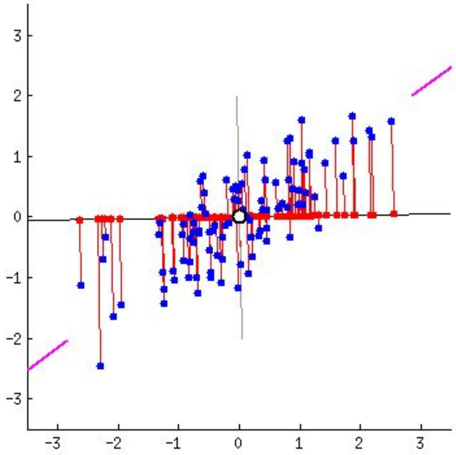


- **Feature extraction**



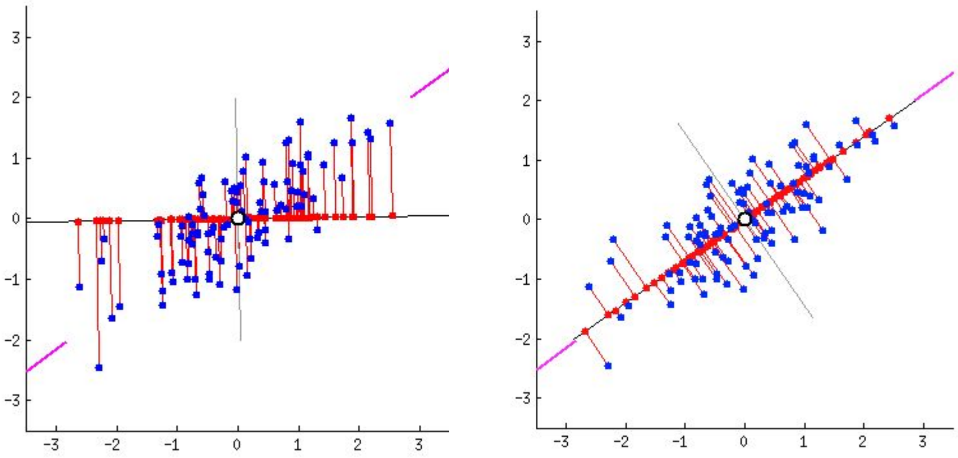
Dimensionality reduction

- **Feature selection**
  - variance feature selection
  - regularization: lasso, ridge, elastic net
- **Feature extraction**
  - PCA
  - tSNE
  - UMAP



Dimensionality reduction

- **Feature selection**
  - variance feature selection
  - regularization: lasso, ridge, elastic net
- **Feature extraction**
  - PCA
  - tSNE
  - UMAP



PC1:

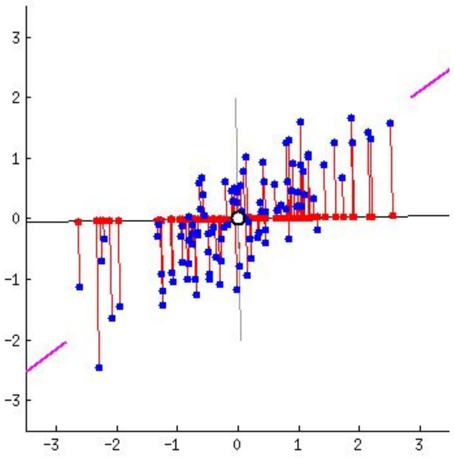
eigenvector ν - direction of variance

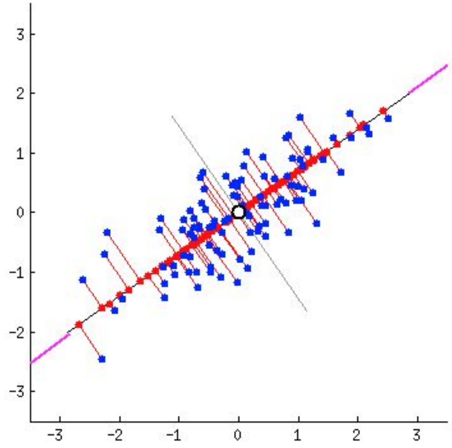
eigenvalue λ - scale of variance

Dimensionality reduction

HANDS-ON

- **Feature selection**
  - variance feature selection
  - regularization: lasso, ridge, elastic net
- **Feature extraction**
  - PCA
  - tSNE
  - UMAP





PC1:

eigenvector ν - direction of variance

eigenvalue λ - scale of variance

Dimensionality reduction

- **Feature selection**

  variance feature selection

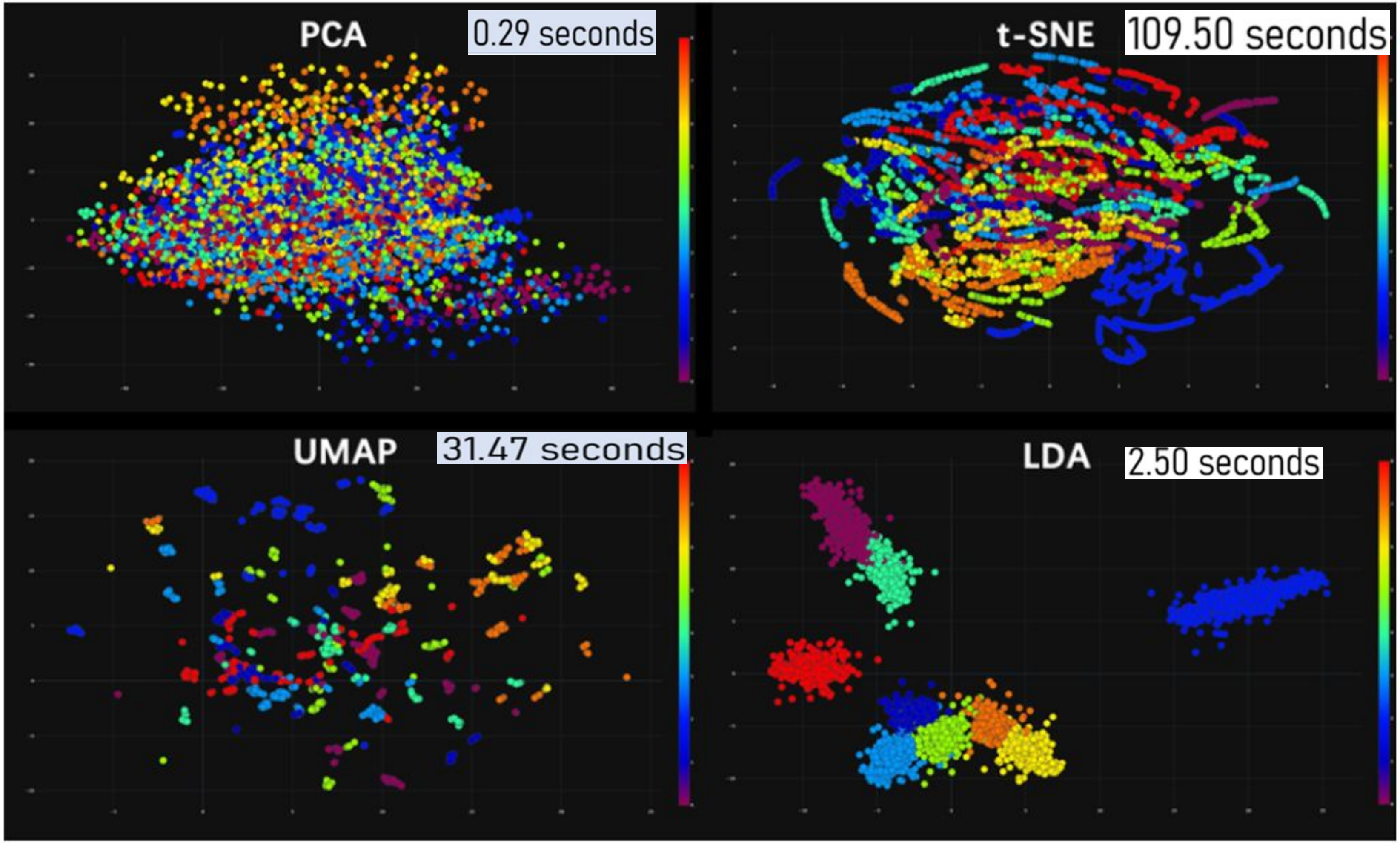
  regularization: lasso, ridge, elastic net

- **Feature extraction**

  PCA

  tSNE

  UMAP


PCA
0.29 seconds
t-SNE
109.50 seconds
UMAP
31.47 seconds
LDA
2.50 seconds

<table>
<tr>
<td>

<u>PCA</u>

linear projection
It finds directions that minimizes the variance in the dataset
Highly influenced by the outliers
It can’t capture the non-linear dependencies

</td>
<td>

<u>t-SNE</u>

Non-linear projection
preserve topology neighbourhood structure
Similar labels are clustered together and does better job than PCA, but it’s just for visualisation. It used PCA for its coordinate calculation.

</td>
</tr>
<tr>
<td>

<u>UMAP</u>

non-linear dimension reduction
It works by modelling the manifolds and assumes that there are manifolds on which data is uniformly distributed, underlying manifolds are locally connected.
it is very effective for visualizing clusters or groups of data points and their relative proximities.

</td>
<td>

<u>LDA</u>

Linear projection
it focuses on maximizing the separability among known categories by creating a new linear axis and projecting the data points on that axis.

</td>
</tr>
</table>