Professur für Audio-Signalverarbeitung
Fakultät für Elektro- und Informationstechnik
Technische Universität München

Unsupervised Classification of vibro-acoustic signals based on machine learning

Forschungspraxis

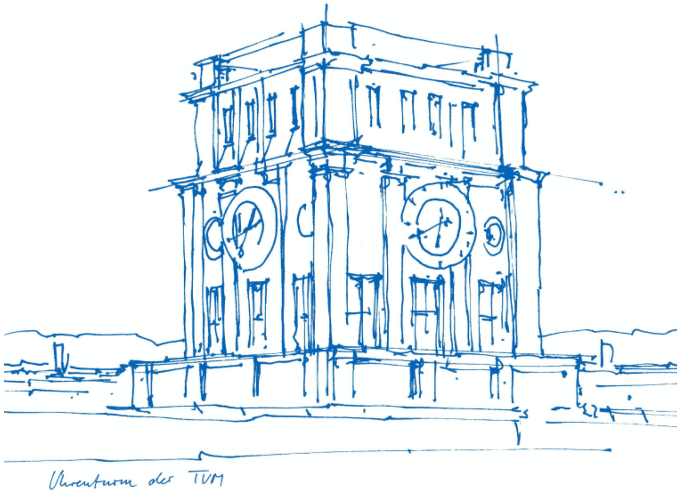
Zhen Zhou

Supervisor :

Norbert Kolotzek

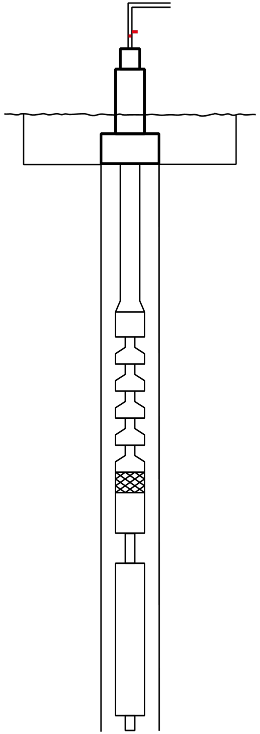
Prof. Dr.-Ing. B. Seeber

Munich, 9. February 2021



Motivations

- 44 spectral & temporal features of measurements (Dec. 2019 - Apr. 2020)
- Currently, features labeled manually



1) **Can we classify the vibro-acoustic signals based on unsupervised learning?**
2) **Can we extract information about the machine state or abnormal behavior from the measurements?**

Sketch from: Kolotzek N., Seeber B.U., Baumann, H. (2019) Obertägiges, akustisches Monitoring im hörbaren Frequenzbereich einer Tiefkreiselpumpe, Der Geothermie Kongress 2019, München

Main idea

- 1. Data Preprocessing for Artificial Neural Networks

- 2. AutoEncoder - ANN
  - Self-Supervised Learning - Not dependent on labels
  - Implement data compression

- 3. Clustering algorithm – Unsupervised Learning
  - DBSCAN
  - K-Means
  - (GMM – Actually not performing well in this case)

Data Preprocessing

- Data structure
  - 1 recording lasts 8 minutes every 90 Minutes
  - Each file split into 319 blocks. (3 sec with 50% overlap)
  - Extraction of 44 features in each block
  - Total of 559845 data analyzed.

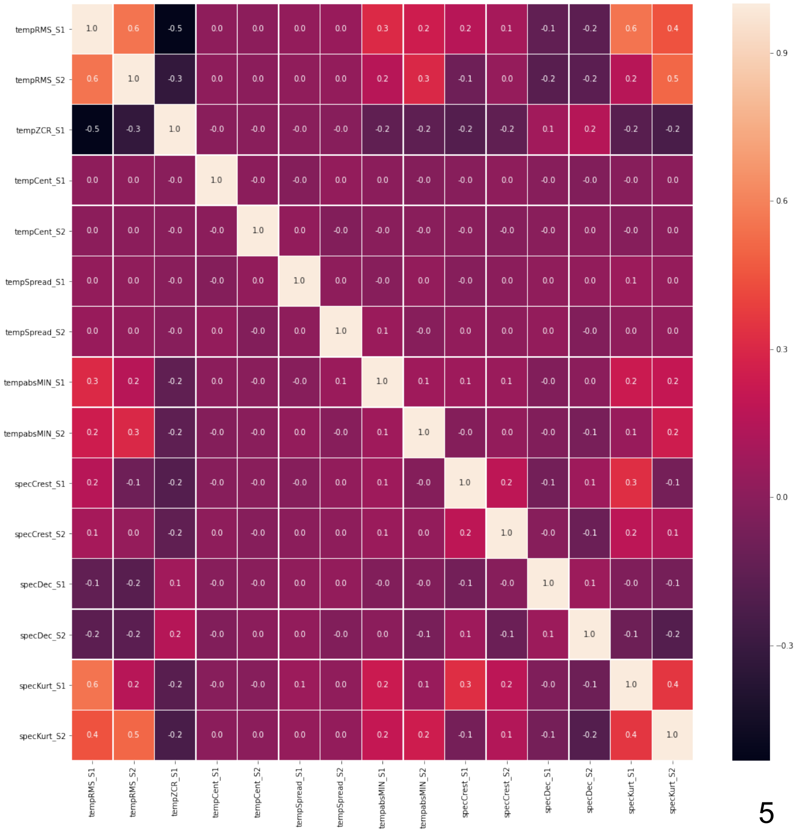
- Data reduction
  - Because of redundant information & noise
  - ➢ Decide to select randomly 1 block in each recording
  - ➢ Current dimension: (1755, 44)

Features Selection

- Compare the correlation of the data
  - 0.8<=|r|<1: high correlation;
  - 0.3<=|r|<0.8: middle correlation
  - ➢ Here threshold = 0.6
    (retain the appropriate number of columns)

➢ Current dimension: (1755, 15)

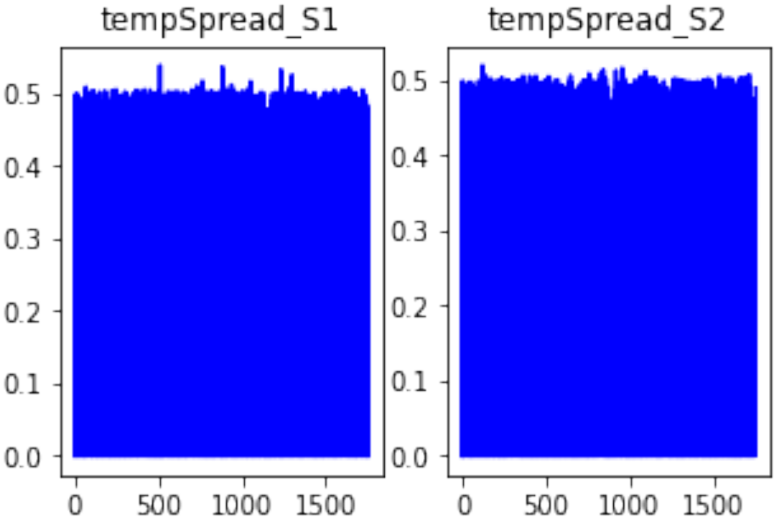
- The correlation heatmap as shown -->



	tempRMS_S1	tempRMS_S2	tempZCR_S1	tempCent_S1	tempCent_S2	tempSpread_S1	tempSpread_S2	tempabsMIN_S1	tempabsMIN_S2	specCrest_S1	specCrest_S2	specDec_S1	specDec_S2	specKurt_S1	specKurt_S2
tempRMS_S1	1.0	0.6	-0.5	0.0	0.0	0.0	0.0	0.3	0.2	0.2	0.1	-0.1	-0.2	0.6	0.4
tempRMS_S2	0.6	1.0	-0.3	0.0	0.0	0.0	0.0	0.2	0.3	-0.1	0.0	-0.2	-0.2	0.2	0.5
tempZCR_S1	-0.5	-0.3	1.0	-0.0	-0.0	-0.0	-0.0	-0.2	-0.2	-0.2	-0.2	0.1	0.2	-0.2	-0.2
tempCent_S1	0.0	0.0	-0.0	1.0	-0.0	-0.0	0.0	0.0	-0.0	-0.0	0.0	-0.0	-0.0	-0.0	0.0
tempCent_S2	0.0	0.0	-0.0	-0.0	1.0	0.0	-0.0	-0.0	-0.0	-0.0	-0.0	-0.0	0.0	-0.0	0.0
tempSpread_S1	0.0	0.0	-0.0	-0.0	0.0	1.0	0.0	-0.0	0.0	-0.0	0.0	0.0	0.0	0.1	0.0
tempSpread_S2	0.0	0.0	-0.0	0.0	-0.0	0.0	1.0	0.1	-0.0	0.0	0.0	0.0	-0.0	0.0	0.0
tempabsMIN_S1	0.3	0.2	-0.2	0.0	-0.0	-0.0	0.1	1.0	0.1	0.1	0.1	-0.0	0.0	0.2	0.2
tempabsMIN_S2	0.2	0.3	-0.2	-0.0	-0.0	0.0	-0.0	0.1	1.0	-0.0	0.0	-0.0	-0.1	0.1	0.2
specCrest_S1	0.2	-0.1	-0.2	-0.0	-0.0	-0.0	0.0	0.1	-0.0	1.0	0.2	-0.1	0.1	0.3	-0.1
specCrest_S2	0.1	0.0	-0.2	0.0	-0.0	0.0	0.0	0.1	0.0	0.2	1.0	-0.0	-0.1	0.2	0.1
specDec_S1	-0.1	-0.2	0.1	-0.0	-0.0	0.0	0.0	-0.0	-0.0	-0.1	-0.0	1.0	0.1	-0.0	-0.1
specDec_S2	-0.2	-0.2	0.2	-0.0	0.0	0.0	-0.0	0.0	-0.1	0.1	-0.1	0.1	1.0	-0.1	-0.2
specKurt_S1	0.6	0.2	-0.2	-0.0	-0.0	0.1	0.0	0.2	0.1	0.3	0.2	-0.0	-0.1	1.0	0.4
specKurt_S2	0.4	0.5	-0.2	0.0	0.0	0.0	0.0	0.2	0.2	-0.1	0.1	-0.1	-0.2	0.4	1.0

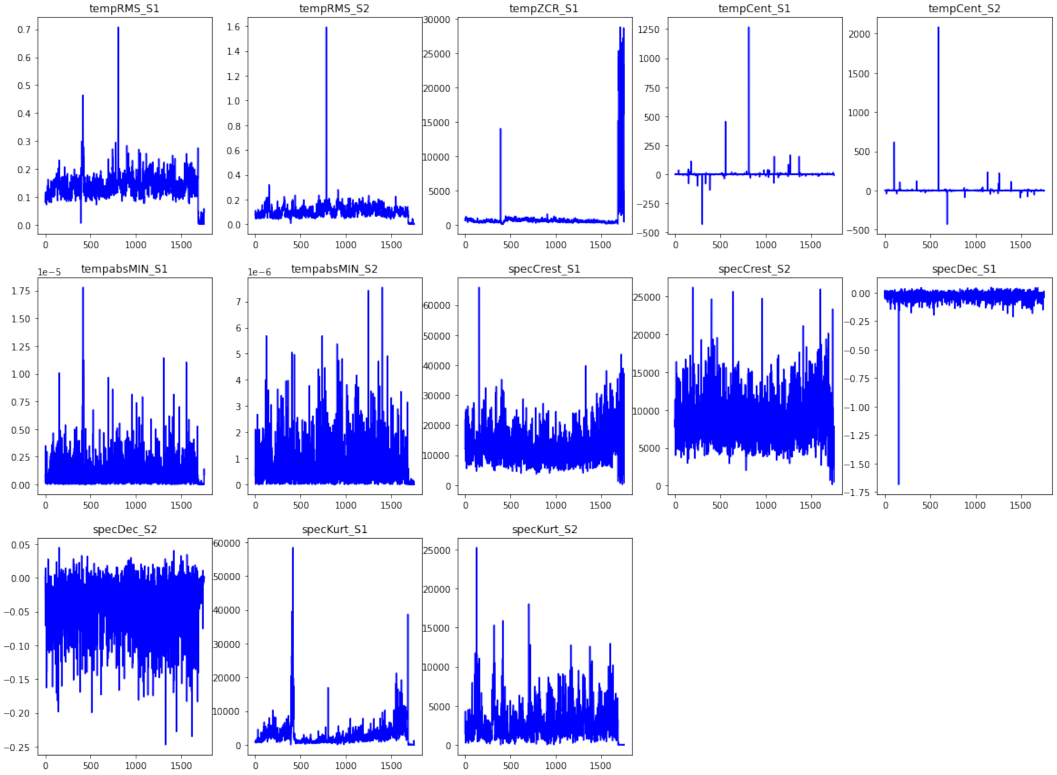
Features Selection

- Delete noisy features
  - According to the data visualization, the following two columns will be deleted.



Features Selection

- Remained features ->



- Current dimension:

  (1755, 13)

Data Preprocessing for ANN

- PCA - Principal component analysis
  - Choose first 10 components, which can express 94% of the variance.
  - Function: Reduce the dimensions; Remove the noise; Remove redundant information

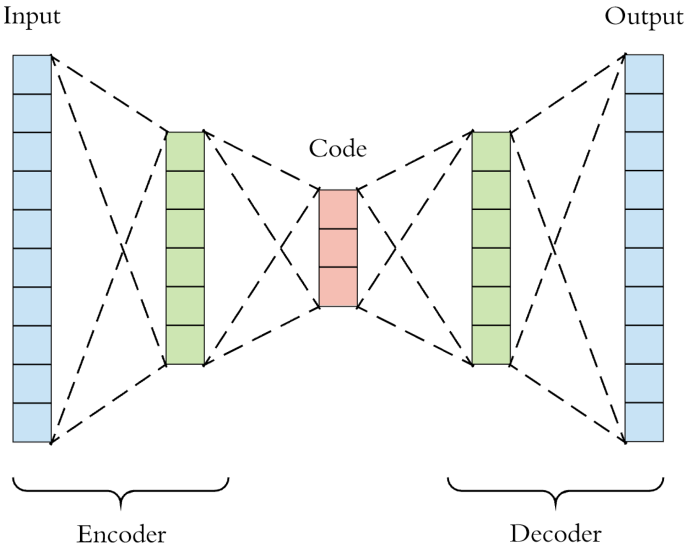
- Data standardization
  - Linearly map each dimension feature to the specified interval, [0, 1]
    - The value range of ReLU activation function is [0, 1]

➢ Current dimension: (1755, 10)

AutoEncoder

- Explanation: Encoder & Decoder
  - Self-Supervised Learning - Not dependent on labels
  - Implement data compression
- Application:
  - Dimensionality reduction
  - Anomaly detection
- Only choose ‘Code’ Area
- Realize unsupervised dimensionality

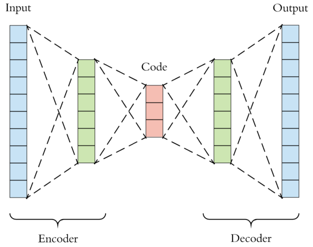
  reduction



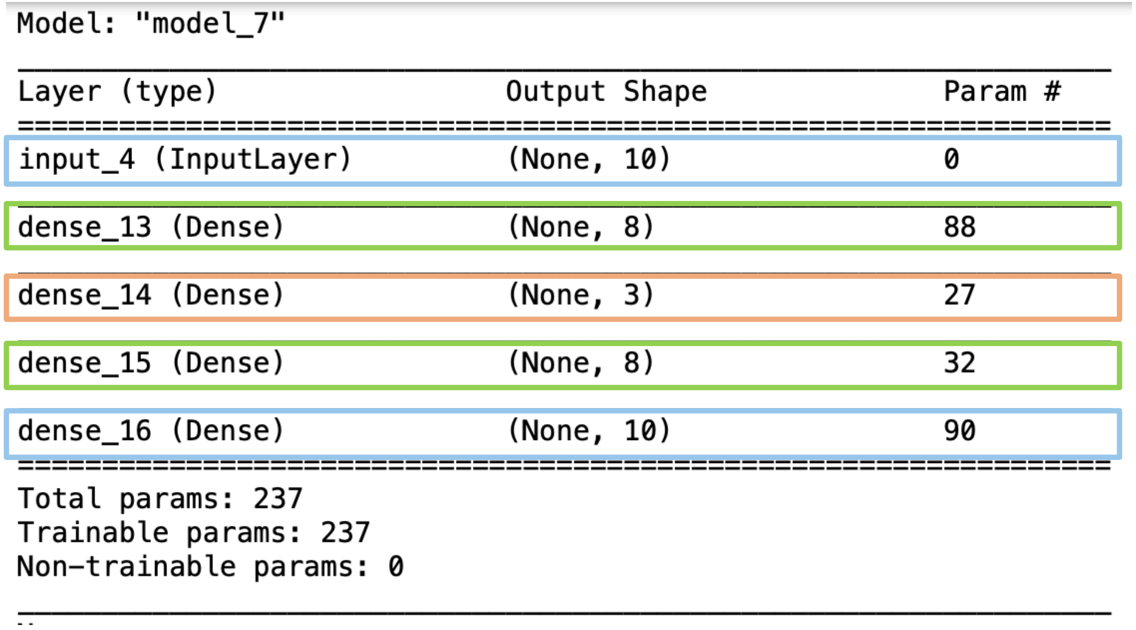
Taken from: https://towardsdatascience.com/generating-images-with-autoencoders-77fd3a8dd368 (checked on: 05.02.2021)

Professur für Audio-Signalverarbeitung
Fakultät für Elektro- und Informationstechnik
Technische Universität München

AutoEncoder



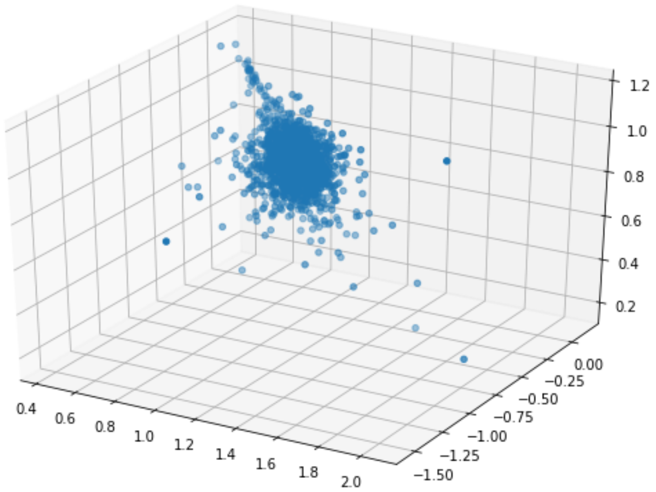
- Encoder:
  - input_4: (, 10)
  - dense_13: (, 8)
- Code:
  - dense_14: (, 3)
- Decoder:
  - dense_15
  - dense_16
- Current dimension: (1755, 3)



```
Model: "model_7"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_4 (InputLayer)         (None, 10)                0
_________________________________________________________________
dense_13 (Dense)             (None, 8)                 88
_________________________________________________________________
dense_14 (Dense)             (None, 3)                 27
_________________________________________________________________
dense_15 (Dense)             (None, 8)                 32
_________________________________________________________________
dense_16 (Dense)             (None, 10)                90
=================================================================
Total params: 237
Trainable params: 237
Non-trainable params: 0
_________________________________________________________________
```

AutoEncoder

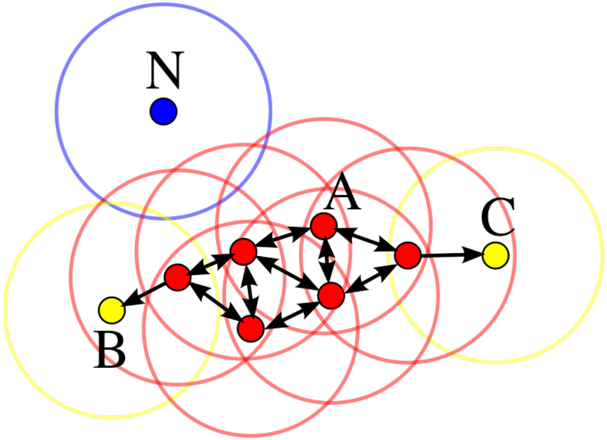
- Output from Encoder:
  - (1755, 10) --> (1755, 3)
  - Distribution of reduced data



Clustering - DBSCAN

- Explanation
  - Density-based clustering algorithm
  - Divide areas with sufficiently high density into clusters
  - Performs well in noisy spaces

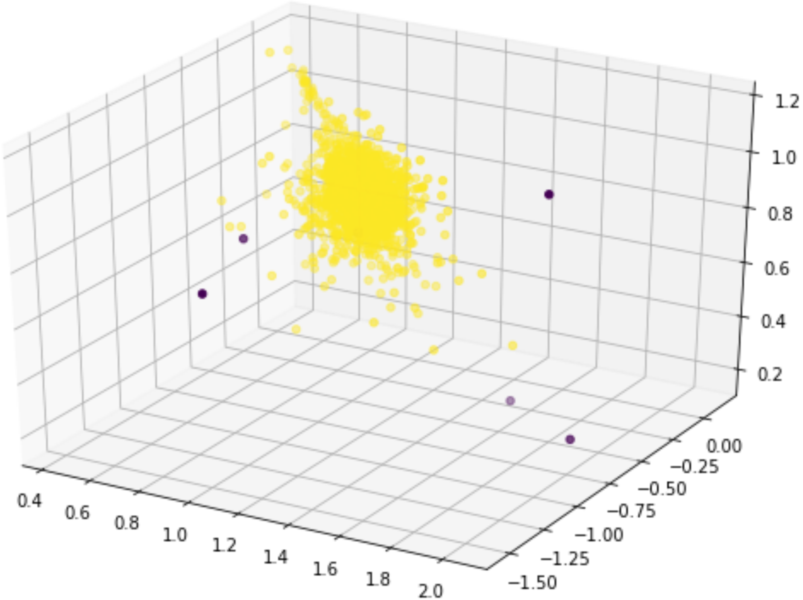
- Algorithm Description:
  - Input: database containing n objects, radius e, minimum number MinPts;
  - Output: All generated clusters meet the density requirement.



Taken from: https://en.wikipedia.org/wiki/DBSCAN (checked on: 05.02.2021)

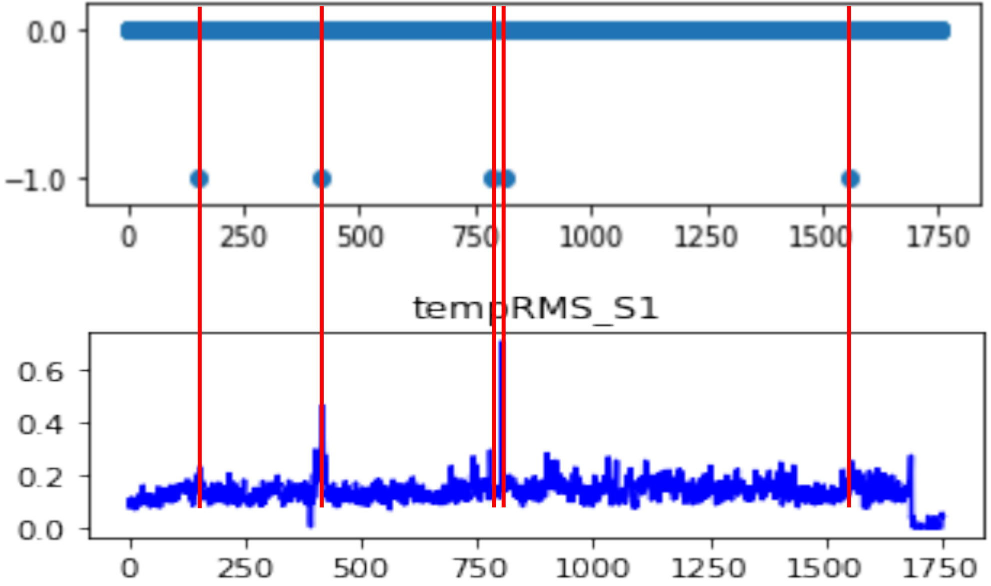
Clustering - DBSCAN

- Parameters: radius e=0.3; MinPts = 5



Clustering - DBSCAN

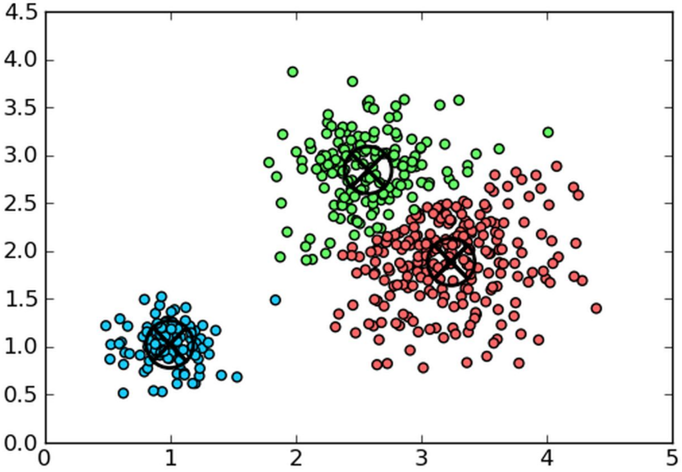
- Plot the result for classification
  - Label -1: abnormal behavior
- The bottom row corresponds to the original data index
- Only observe abnormal situations
- But no temporal in/decrease for heath status



Clustering - KMeans

- Explanation
  - Distance-based clustering algorithm
  - Fast calculation speed

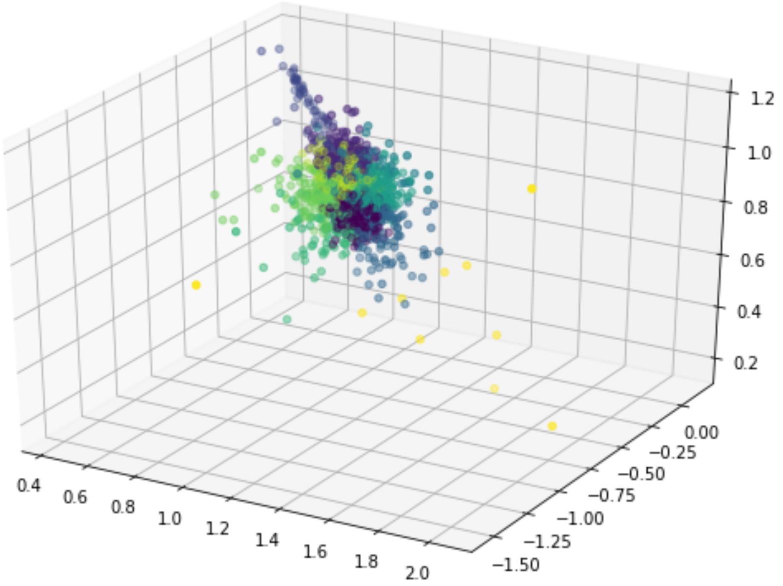
- Algorithm Description:
  - According to a certain distance function repeatedly divide the data into k clusters
  - Need to specify the number of clusters



Taken from: https://zhuanlan.zhihu.com/p/37875887 (checked on: 05.02.2021)

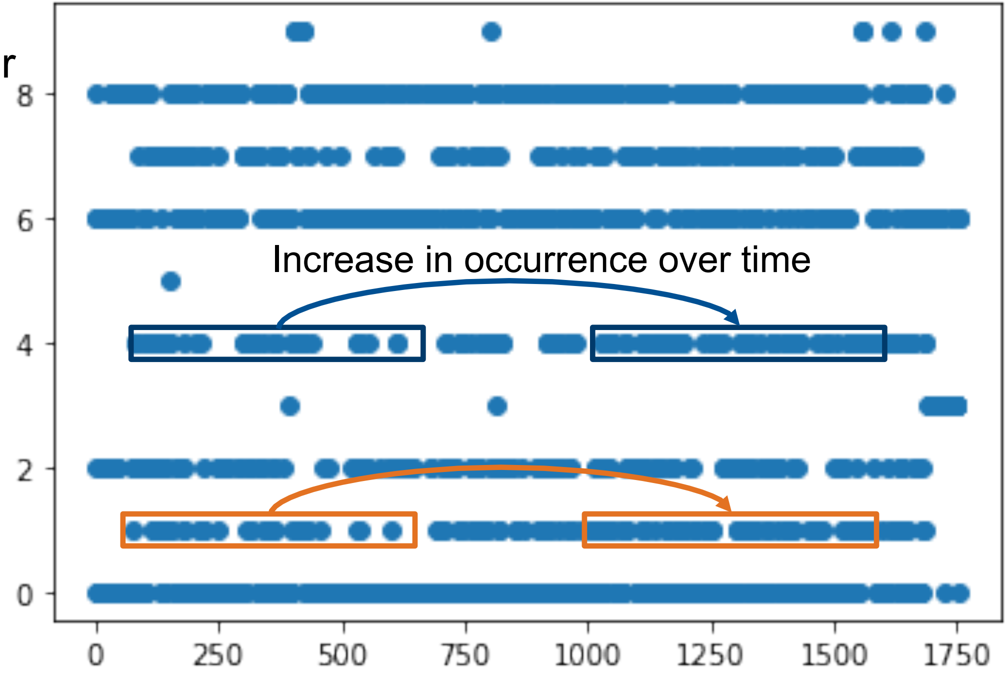
Clustering - KMeans

- Parameters: Num_cluster=10



Clustering - KMeans

- Plot the result for classification
  - Some labels: abnormal behavior
  - Others need further analysis

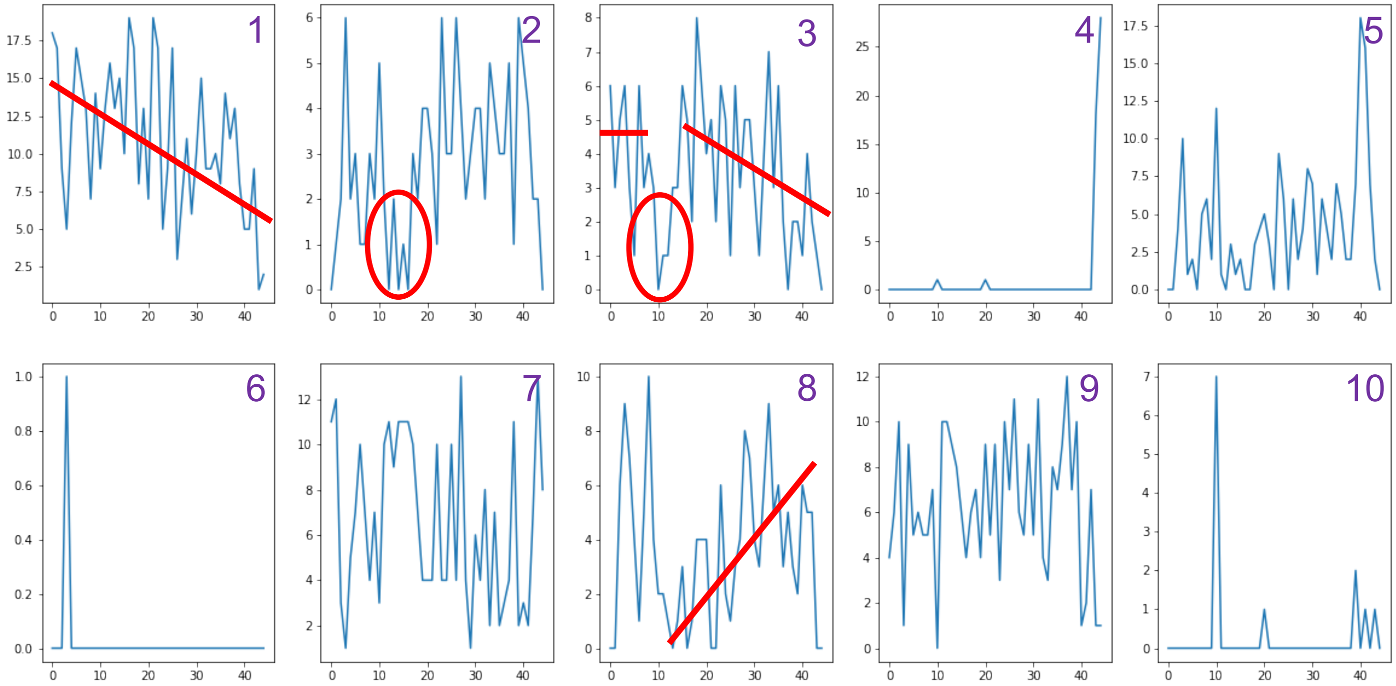


Increase in occurrence over time

Professur für Audio-Signalverarbeitung
Fakultät für Elektro- und Informationstechnik
Technische Universität München

Clustering - KMeans

- Plot: How often are the clusters chosen over time?
- Health state
- Abnormal



Summary / Conclusion

- Transforming wealth of features (high dimensionality) to meaningful low dimensionality

- Detecting critical / abnormal behavior

- After clustering increase or decrease on occurrence could be used as a prediction for health status