

Date

NO.

Ex. Classification vs Regression

- 둘 다 지도 학습의 방법

Classification	→	Regression
• 목표 : 이진, 다중 클래스 (예: 스팸/정상, 양성/음성)		• 목표 : 실수 (예: 가격, 점수)
• 변수 : 범주, 수치		• 변수 : 수치
• 결과 : 클래스, 확률, 점수		• 결과 : 값
• 예시 : 스팸/정상, 양성/음성, 추천		• 예시 : 가격, 점수, 점수

• Classification → 범주 예측
Regression → 수치 예측

6. 머신러닝에서 특징의 선택 (Feature Selection)

- 데이터에서 불필요한 특징 (Feature)을 제거하여 모델의 성능을 향상시키는 것

↳ 불필요한 특징을 제거하면 모델의 성능을 향상시킬 수 있다.

- 특징 선택의 필요성

- 모델의 성능 : 불필요한 특징이 포함되면 모델의 성능이 저하된다.
- 계산 비용 : 불필요한 특징을 포함하면 계산 비용이 증가한다.
- 모델의 해석 : 불필요한 특징을 포함하면 모델의 해석이 어렵다.

↳ 특징 선택 (Feature Selection) 방법

- 필터 방법 : 특징의 분포, 분산 등을 기준으로 특징을 선택한다.
- 래퍼 방법 : 모델의 성능을 기준으로 특징을 선택한다.
- 임베딩 방법 : 특징을 임베딩하여 특징을 선택한다.

7. Dimensionality Reduction

1. 차원 축소 (Dimensionality Reduction)

2. 차원 축소 방법

1. PCA : 주성분 분석을 통해 차원을 축소한다.
2. LDA : 선형 판별 분석을 통해 차원을 축소한다.
3. t-SNE : t-분포를 사용하여 차원을 축소한다.
4. Autoencoder : 오토인코더를 사용하여 차원을 축소한다.
5. Variational Autoencoder : 변형 오토인코더를 사용하여 차원을 축소한다.

8. Prüfung da Leistung auswerten ist erforderlich. (Prüfung, Wert, Bewertung)

o Prüfung

- Prüfung ist ein Verfahren zur Bestimmung der Leistung einer Person oder Gruppe von Personen.
- Ziel: Leistungs feststellung und Beurteilung des Lernfortschritts.
- Prüfung ist ein Verfahren zur Bestimmung der Leistung einer Person oder Gruppe von Personen.
- Prüfung ist ein Verfahren zur Bestimmung der Leistung einer Person oder Gruppe von Personen.

o Prüfung

- | | | |
|---|----------------|----------------|
| <u>Prüfung</u> | <u>Prüfung</u> | <u>Prüfung</u> |
| - <u>Prüfung</u> <u>ist</u> <u>ein</u> <u>Verfahren</u> <u>zur</u> <u>Bestimmung</u> <u>der</u> <u>Leistung</u> <u>einer</u> <u>Person</u> <u>oder</u> <u>Gruppe</u> <u>von</u> <u>Personen</u> . | <u>Prüfung</u> | <u>Prüfung</u> |
| - <u>Prüfung</u> <u>ist</u> <u>ein</u> <u>Verfahren</u> <u>zur</u> <u>Bestimmung</u> <u>der</u> <u>Leistung</u> <u>einer</u> <u>Person</u> <u>oder</u> <u>Gruppe</u> <u>von</u> <u>Personen</u> . | <u>Prüfung</u> | <u>Prüfung</u> |
| - <u>Prüfung</u> <u>ist</u> <u>ein</u> <u>Verfahren</u> <u>zur</u> <u>Bestimmung</u> <u>der</u> <u>Leistung</u> <u>einer</u> <u>Person</u> <u>oder</u> <u>Gruppe</u> <u>von</u> <u>Personen</u> . | <u>Prüfung</u> | <u>Prüfung</u> |
| - <u>Prüfung</u> <u>ist</u> <u>ein</u> <u>Verfahren</u> <u>zur</u> <u>Bestimmung</u> <u>der</u> <u>Leistung</u> <u>einer</u> <u>Person</u> <u>oder</u> <u>Gruppe</u> <u>von</u> <u>Personen</u> . | <u>Prüfung</u> | <u>Prüfung</u> |

9. Prüfung ist ein Verfahren zur Bestimmung der Leistung einer Person oder Gruppe von Personen.

- | | | |
|---|----------------|----------------|
| <u>Prüfung</u> | <u>Prüfung</u> | <u>Prüfung</u> |
| - <u>Prüfung</u> <u>ist</u> <u>ein</u> <u>Verfahren</u> <u>zur</u> <u>Bestimmung</u> <u>der</u> <u>Leistung</u> <u>einer</u> <u>Person</u> <u>oder</u> <u>Gruppe</u> <u>von</u> <u>Personen</u> . | <u>Prüfung</u> | <u>Prüfung</u> |
| - <u>Prüfung</u> <u>ist</u> <u>ein</u> <u>Verfahren</u> <u>zur</u> <u>Bestimmung</u> <u>der</u> <u>Leistung</u> <u>einer</u> <u>Person</u> <u>oder</u> <u>Gruppe</u> <u>von</u> <u>Personen</u> . | <u>Prüfung</u> | <u>Prüfung</u> |
| - <u>Prüfung</u> <u>ist</u> <u>ein</u> <u>Verfahren</u> <u>zur</u> <u>Bestimmung</u> <u>der</u> <u>Leistung</u> <u>einer</u> <u>Person</u> <u>oder</u> <u>Gruppe</u> <u>von</u> <u>Personen</u> . | <u>Prüfung</u> | <u>Prüfung</u> |
| - <u>Prüfung</u> <u>ist</u> <u>ein</u> <u>Verfahren</u> <u>zur</u> <u>Bestimmung</u> <u>der</u> <u>Leistung</u> <u>einer</u> <u>Person</u> <u>oder</u> <u>Gruppe</u> <u>von</u> <u>Personen</u> . | <u>Prüfung</u> | <u>Prüfung</u> |

10. Feature Engineering vs Feature Selection

	Feature Engineering	vs	Feature Selection
목적	새로운 특징 추출하기 또는 기존 특징		불필요한 특징 제거 또는 특징/변수 선택
주요 방법	수치, 범주, 시계열 등 다양한 변형		필터, 래퍼, 임베딩 등
결과	새로운 특징 추가		변수 개수 감소
목적	모델 성능 향상시키기 위한 것		모델 복잡도 낮추기
활용 방법	새로운 특징 추출하기		모든 특징을 사용하거나 선택

11. Feature Engineering 방법론 (4가지, 이진화, 인코딩)

- 이진화

- 이진화 방법: 범주형 변수를 0/1로 변환하여 모델에 입력
- 이진화 방법: 범주형 변수를 0/1로 변환하여 모델에 입력
- 이진화 방법: 범주형 변수를 0/1로 변환하여 모델에 입력
- 이진화 방법: 범주형 변수를 0/1로 변환하여 모델에 입력

- 인코딩

Label → Label Encoder, One-hot Encoder, Embedding

Ordinal → Ordinal Encoder, Label Encoder

Matrix Vector → Matrix Vector, Matrix Vector, Matrix Vector

Matrix Vector → Matrix Vector, Matrix Vector, Matrix Vector

12. EPA (Exploring Data Analysis)란? 데이터 분석 과정 (보통, 상식적인)

- 탐색적 데이터 분석, 이산치, 연속치, 범주형, 양적치에 따라 어떤 분석 방법을 이용하느냐?
- 데이터 분석 과정
 - 문제 인식
 - 목표: 어떤 데이터를 분석할 것인가?
 - 방법: 탐색적, 진단적, 예측 → 모든 상황
모형(모델), 추정(estimate), 회귀분석
 - 유의성 검정
 - 독립: 두 변수 간 상관관계가 있는지 확인
 - 양적: 두 변수 간 상관관계가 있는지, 유의성 확인

13. 회귀분석: 각 변수가 얼마나 영향을 미치는가? 회귀분석의 중요성?

회귀 - 선형회귀 $y = Ax + b$ 에서 b 값, 독립변수 x 의 0일 때 y 값

기분 - x 값이 1일 때, 독립변수 x 의 1일 때 y 값, y 값의 변화량을 나타냄.

회귀분석의 의미: 회귀분석 기법으로 회귀(변환)한 b 값과 x 값의 관계를 분석함.
이를 통해 회귀분석을 이용하여 예측할 수 있다.

회귀(변환): 기온이 높아지면, 일출시간(변환)을 나타냄.

회귀분석의 의미: 회귀분석(변환)을 사용하여 예측할 수 있다.

회귀분석의 의미: 회귀분석을 이용하여 예측할 수 있다.

회귀(변환): 기온이 높아지면, 일출시간(변환)을 나타냄.

회귀분석의 의미: 회귀분석을 이용하여 예측할 수 있다.

26. 전이도(Transit) 분석 (Transit) - 지 지 지 (Gini Index)란 무엇인가?

- 전이도(Transit) 분석은 데이터를 분석할 때, 지 지 지 (Gini Index)란 무엇인가?
- 지 지 지 (Gini Index)란 무엇인가?
- 지 지 지 (Gini Index)란 무엇인가?

27. 상관계수란 무엇인가?

- 상관계수란 무엇인가?
- 상관계수란 무엇인가?

30. 부른 드레싱 (buds dressing)이란 무엇인가?

- 주어진 데이터에서 중요한 특징을 추출하여 해당 샘플은 여러 개의 특징을 가진 데이터를 만드는 방법. 주어진 데이터에 의거하여 분류.

ex) $Var: \{A, B, C, D, E\}$

↓
sample: 1. $\{A, C, D, E\}$
2. $\{B, D, C, E, D\}$
3. $\{A, A, B, D\}$

31. 부팅 (Booting)이란 무엇인가?

- 부팅은 시스템을 실행하여 여러 가지 작업을 수행하고, 여러 영역을 검사하여 상태를 항상 같은 상태로 유지한다.

→ 데이터에 해당 샘플을 통해 여러 영역을 독립적으로 학습시키고, 그 결과를 결합하여 모델의 성능을 높이는 것이다.

32. 주성분 분석 (PCA)이란 무엇인가?

- 주어진 데이터에서 ^{특정} 주요한 특징을 추출하여 낮은 차원으로 변환하는 선형 변환 방법이다. 차원을 줄여 데이터에 복잡성을 낮추는 방법.

- 차원 축소 < 노이즈 제거
데이터 선형화