

## <용어체크>

### 머신러닝

신경망의 학습처럼 기계가 학습한다는 개념을 의미하는 것이며 입력 데이터의 특성과 분포 경향 등에서 자동으로 데이터를 나누거나 재구성한다. 통계 기반 머신러닝에서는 확률 개념이 크게 관여한다.

### EM 알고리즘

숨겨진 변수를 포함하는 모델에 사용하는 가상 데이터 전체의 가능도 함수를 바탕으로 불완전한 데이터에서 최대 가능도 추정량을 구하는 방법이다.

### 베이즈 판별 분석

$N$ 가지 종류의 데이터  $x$ 와 추론 대상 매개변수  $i$ 가 있는 모집단분포  $f(x|i)$ 와 사전 분포  $\pi(i)$ 로부터 사후 확률이 최대가 되는 모집단 분포  $f(x|i)$ 의 유래를 판정하는 것이다.

## <학습내용>

### 통계 모델과 확률 분포

### 주요 기저 함수

### 베이즈 통계학

## <학습목표>

확률 기반의 통계 모델을 설명할 수 있다.

통계 모델에 사용되는 기저 함수들을 설명할 수 있다.

베이즈 정리를 설명할 수 있다.

Q. 학습에 관해서는 요즘 머신러닝이 많이 이야기 나오고 있는데 기계 학습에는 어떤 것들이 있을까요?

: 기계 학습은 우리가 인공지능을 하는 목표의 방법이라고 볼 수 있습니다. 과연 기계가 학습을 하기 위해서는 어떻게 해야 할까요? 일단은 기본적으로 학습에 대해서는 데이터에 학습 후 정답을 같이 주어 하는 학습을 볼 수 있습니다. 문제와 답을 동시에 주는 것으로 이것을 지도학습이라고 합니다. 그리고 다음으로는 데이터에 정답을 주지 않는 학습도 있습니다. 그럴 경우는 비지도 학습 또는 자율 학습이라고 합니다. 많은 데이터를 통해 답을 스스로 학습을 하는 것입니다. 다음은 스스로 학습의 방법을 조절하는 학습이 있습니다. 강화 학습이라고 합니다. 이런 3가지 학습법을 이용하여 기계가 사람보다 뛰어난 결과의 내용을 보일 수 있습니다. 앞으로 계속 발전하는 학습법을 지켜 보도록 하세요.

## 통계 모델과 확률 분포

- ▶ 머신 러닝은 기계가 학습을 통해 네트워크의 가중치를 변경하는 것을 말한다.
- ▶ 머신 러닝에서 데이터에 정답 정보가 결합된 학습 데이터로 데이터의 특징을 모델링하는 것을 지도학습이라 한다.
- ▶ 입력 데이터의 정답을 모르는 상태에서 사용하는 것을 자율 학습이라고 한다.
- ▶ 정규 분포 이외의 분포를 다루는 것을 일반화 선형 모델이라고 한다.

## 주요 기저 함수

- ▶ 정규 분포는 가우스 분포라고도 하며 가장 많이 사용하는 분포 개념이다.
- ▶ 주요 기저 함수는 연속 확률 분포와 이산 확률 분포로 구분할 수 있다.
- ▶ 목적 함수와 비슷한 함수로 손실 함수를 사용하며 손실 함수의 최소값을 구할 때 경사 하강법 또는 최대 가능도 방법을 사용한다.

## 베이즈 통계학

- ▶ 베이즈 통계학은 조건부 확률에 관한 법칙인 베이즈 정리를 기본으로 한다.
- ▶ 숨겨진 변수를 포함하는 모델에 사용하는 가상 데이터 전체의 가능도 함수를 바탕으로 불완전한 데이터에서 최대가능도 추정량을 구하는 방법을 EM 알고리즘이라 한다.
- ▶ 베이즈 추론의 예로 베이즈 판별 분석이 있으며 데이터  $x$ 가 어떤 모집단 분포에서 유래했는지 파악하는데 사용된다.