

053 임베딩

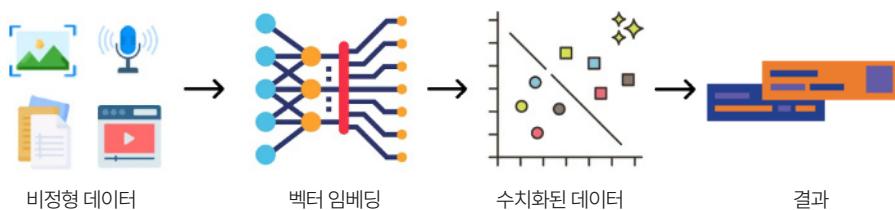
Embedding

비정형 데이터를 수치 벡터로 변환해 의미를 표현하는 기술

- 텍스트·이미지 같은 복잡한 데이터를 숫자로 바꾸어 의미 관계를 계산할 수 있도록 하는 표현 방식
- AI가 단어·개념의 유사성을 이해하고 추론하는 데 사용

임베딩의 개념

임베딩(Embedding)은 텍스트·이미지와 같은 비정형 데이터를 의미 정보를 보존한 채 고차원 벡터 공간으로 변환하는 기술입니다. 이 벡터는 단순한 숫자 변환이 아니라 데이터의 맥락과 의미적 특징을 압축적으로 표현한 형태입니다. 단어나 문장, 이미지와 같은 비정형 데이터는 그대로는 연산이 불가능하기 때문에, 이를 일정한 규칙에 따라 벡터 형태로 변환합니다. 각 벡터는 데이터의 의미적 특징을 반영하고 있어, AI는 이 벡터 간의 거리와 방향을 비교해 유사도나 관계를 계산할 수 있습니다. 예를 들어 '고양이'와 '개'는 서로 가까운 위치에, '고양이'와 '자동차'는 멀리 떨어진 위치에 놓여 의미적 차이를 표현합니다. 이러한 수치화 과정을 통해 AI는 언어나 이미지의 맥락과 의미를 파악할 수 있게 됩니다.



임베딩의 필요성

AI는 단순히 단어의 모양이나 순서만으로는 문맥을 이해하기 어렵습니다. 같은 단어라도 사용되는 상황에 따라 의미가 달라지기 때문입니다. 임베딩은 이러한 한계를 해결해, 데이터의 '의미적 정보'를 보존하면서 수학적으로 처리할 수 있도록 하는 중간 표현층의 역할을 합니다. 이를 통해 AI는 문장의 흐름을 해석하고, 유사한 개념을 묶거나 다른 의미를 구분할 수 있습니다. 예를 들어 "배"가 '신체의 일부'인지 '운송 수단'인지, '과일'인지 구별하려면, 주변 단어와의 관계를 고려해야 하는데 임베딩은 바로 이 문맥 정보를 반영합니다. 또한 텍스트뿐 아니라 이미지·음성 등 다양한 데이터를 동일한 벡터 공간으로 표현할 수 있어, 서로 다른 정보 간 연관성을 분석하는 멀티모달 AI 구현의 기반이 됩니다. 임베딩은 결국 AI가 '이해'와 '추론'을 가능하게 하는 핵심 표현 기술로, 현대 AI 모델의 필수 구성 요소로 자리 잡고 있습니다.