

12. 세이지메이커 Ground Truth

1강. Ground Truth란?

학습목표

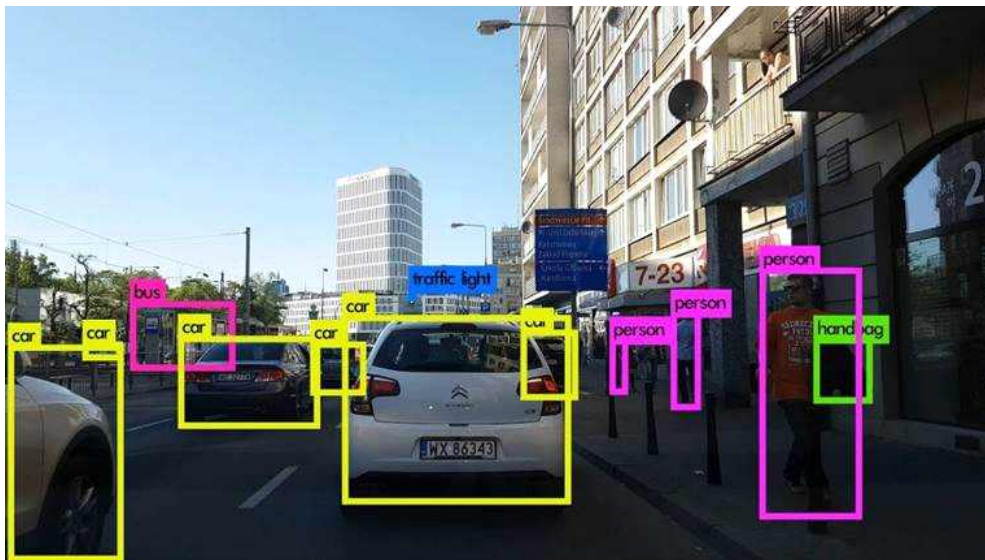
- 세이지메이커 Ground Truth 개요, 이점, 기능에 대해 설명할 수 있다.

학습내용

- 세이지메이커 Ground Truth 개요
- 세이지메이커 Ground Truth 이점
- 세이지메이커 Ground Truth 기능

■ 세상을 잇(IT)다!

- 데이터 레이블링
 - ✓ 데이터 레이블링은 머신러닝이나 딥러닝 모델링 작업 전에 학습 데이터(이미지, 텍스트 파일, 비디오 등)에 특정 값을 부여하는 것



[출처] <https://towardsdatascience.com/computer-vision-for-tracking-8220759eee85>

- 딥러닝, 머신러닝을 위해서는 많은 양의 데이터가 필요

- 좋은 훈련 결과를 위하여 최소 수천에서 수만 개의 데이터의 레이블링을 사람이 직접 작업



- 해결하고자 하는 문제가 복잡할 수록 데이터셋도 점점 복잡해지고 많은 시간 및 비용이 소요됨

- 한 사람이 수천/수만 개의 데이터를 분류하는 것도 어렵지만 여러 사람이 일관된 기준으로 레이블링 작업을 하는 것도 어려움이 따름

✓ 파란색의 기준은 어디까지 일까?



- 데이터 레이블링 작업에는 일관된 기준이 필요함

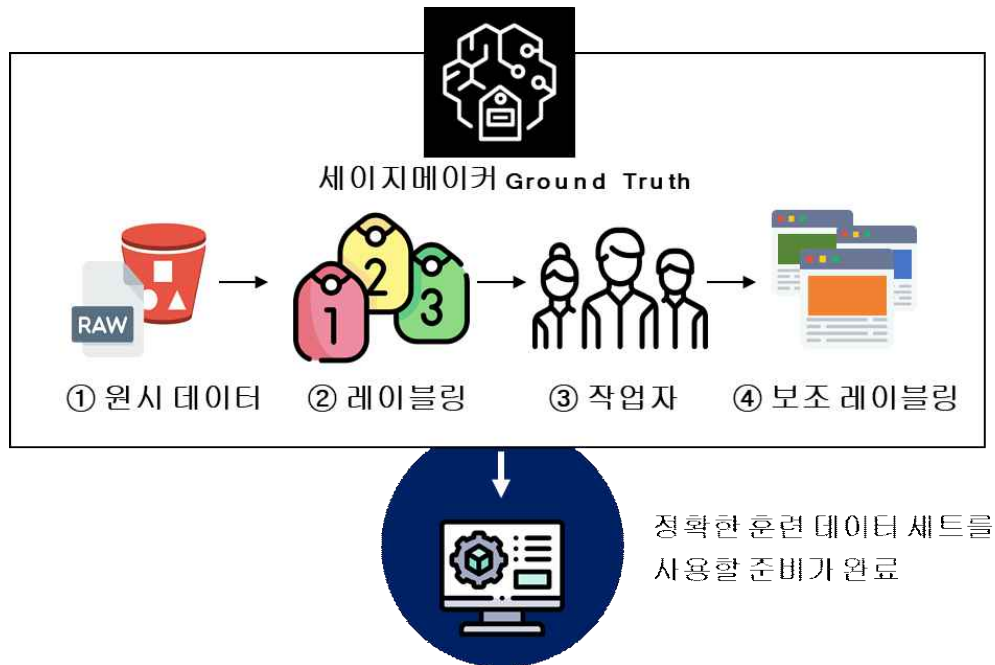
1. 세이지메이커 Ground Truth 개요

- 세이지메이커 Ground Truth : 기계학습을 위한 고도로 정확한 훈련 데이터 세트를 손쉽게 구축할 수 있게 해주는 완전관리형 데이터 레이블링 서비스



[출처] AWS

- 세이지메이커 Ground Truth 작동방식



[출처] AWS

2. 세이지메이커 Ground Truth 이점



- 데이터 레이블 정확도 개선
 - ✓ 세이지메이커 Ground Truth는 주석 통합 및 감사 워크플로우를 통해 레이블의 품질을 개선하도록 지원
 - ✓ 주석 통합은 둘 이상의 데이터 작업자로부터 레이블 입력을 수집하고 이를 결합하여 기계 학습 모델을 위한 단일 데이터 레이블을 생성하는 프로세스
 - ✓ 작업자는 기본 제공 감사 및 검토 워크플로우를 통해 레이블을 검증하고 필요에 따라 조정하여 정확도를 높일 수 있음

- 사용 편의성
 - ✓ 세이지메이커 Ground Truth는 ‘자동 세그먼트’, ‘자동 3D 입방체 스넵핑’, ‘2D 동영상 프레임을 통한 센서 퓨전’과 같은 자동화된 레이블링 기능을 직관적으로 사용자 인터페이스를 통해 제공
 - ✓ 품질 개선과 동시에 데이터 레이블링 태스크에 소요되는 시간을 줄일 수 있음
 - ✓ 데이터 레이블링 작업자는 자동 세그먼트 기능을 사용하여 수 백 번이 아닌 4 번의 클릭으로 객체를 캡처 할 수 있음
- 비용 절감
 - ✓ 세이지메이커 Ground Truth는 데이터 레이블링을 자동으로 수행
 - ✓ 활성 학습 모델을 사용하여 데이터가 레이블링 되고, 모델이 높은 신뢰도로 레이블링 할 수 없는 경우에만 데이터가 사람에게 라우팅 됨
 - ✓ 그 후 사람이 레이블링한 데이터로 기계 학습 모델을 훈련시켜 정확성을 개선
 - ✓ 그 결과, 다음 번 레이블에서는 사람에게 전송되는 데이터가 줄어들어 데이터 레이블링 비용이 최대 70%까지 절감
- 인력 선택
 - ✓ 세이지메이커 Ground Truth에서는 조직 안팎의 작업자와 함께 협력할 수 있는 옵션을 제공
 - ✓ 레이블링 작업을 자체 작업자에게 손쉽게 전송할 수도 있고, 이미 Amazon Mechanical Turk를 통해 머신러닝 관련 작업을 수행하고 있는 500,000명 이상의 독립 계약업체 인력도 이용할 수 있음
 - ✓ 데이터 기밀성 또는 특수한 기술이 요구되는 경우 AWS에서 사전 선별한 공급 업체를 이용하여 품질 및 보안 절차를 수행할 수 있음

3. 세이지메이커 Ground Truth 주요 기능

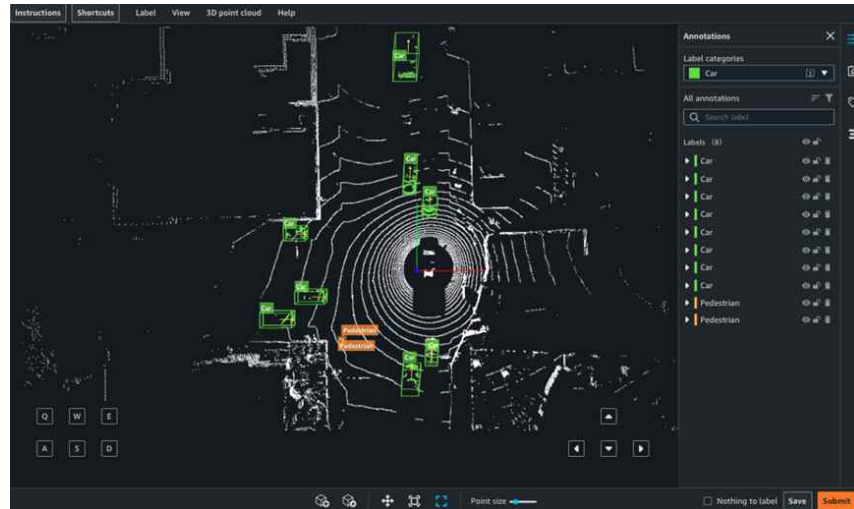


- 3D 포인트 클라우드

- ✓ 3차원(3D) 포인트 클라우드는 단일 시점의 물리적 공간을 3D로 파악하기 위해 빔 탐지 및 거리 측정(LIDAR) 디바이스를 통해 가장 많이 캡처됨

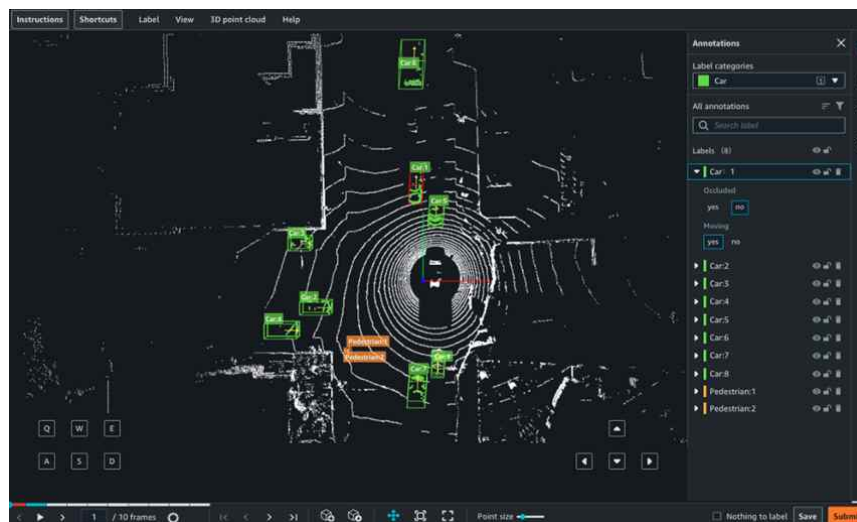
① 객체 탐지

- 포인트 클라우드 내에서 대상 객체의 레이블을 식별할 수 있음



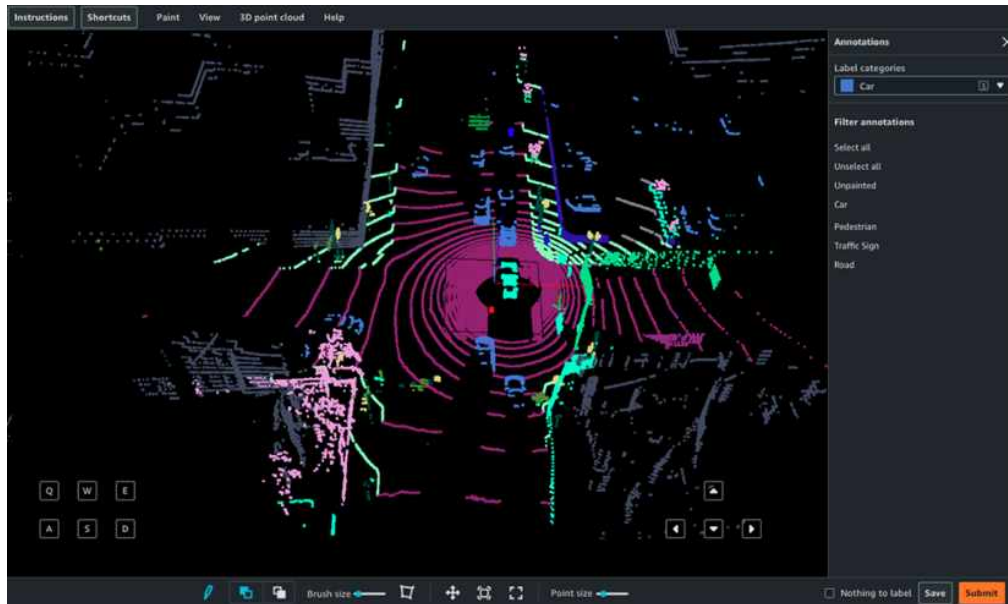
② 객체 추적

- 객체 추적 워크플로에서는 대상 객체의 궤적을 추적할 수 있음



③ 의미 체계 세분화

- 의미 체계 세분화 워크플로에서는 3D 포인트 클라우드의 포인트를 미리 지정한 카테고리로 세분화할 수 있음



• 동영상

- ✓ 기본 제공 워크플로우를 통해 일반적인 비디오 레이블링을 지원

① 비디오 객체

- 비디오 객체 탐지 워크플로우를 활용하면 비디오 프레임의 시퀀스 내에서 원하는 객체를 식별할 수 있음



② 비디오 객체 추적

- 비디오 객체 추적 워크플로우를 활용하면 비디오 프레임의 시퀀스에 걸쳐 원하는 객체를 추적



③ 비디오 클립 분류

- 비디오 클립 워크플로우를 사용하면 비디오 파일을 미리 지정된 범주로 분류할 수 있음



- 이미지

- ✓ 이미지 데이터에 대한 레이블링 워크플로우를 기본적으로 제공

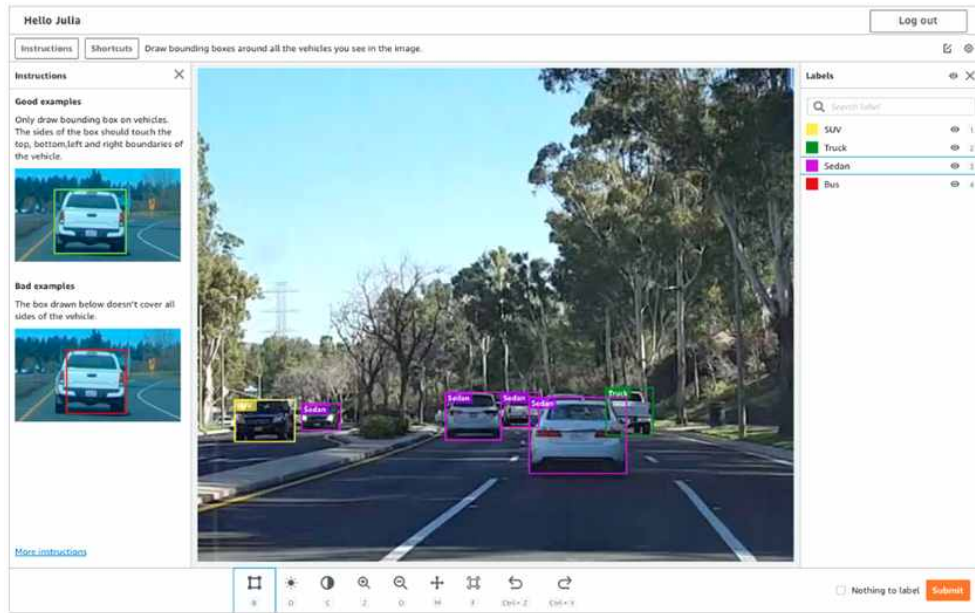
① 이미지 분류

- 이미지 분류는 이미지의 실세계 표현을 토대로 이미지를 식별하는 프로세스
- 사전 정의된 레이블 세트를 기준으로 이미지를 분류하는 과정이 포함
- 이미지 분류는 이미지 전체 컨텍스트를 고려해야 하는 장면 탐지 모델에 유용함



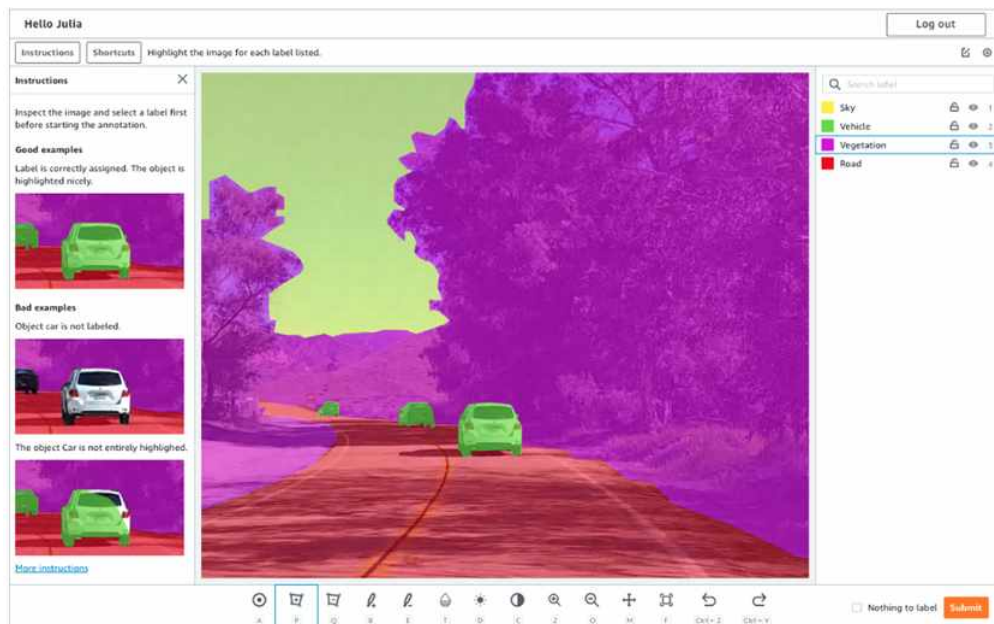
② 객체 탐지

- 객체 탐지 워크플로우를 활용하면 이미지에서 대상 객체(차량, 보행자, 개, 고양이 등)를 식별하고 레이블링할 수 있음
- 레이블링 태스크에는 이미지 내의 대상 객체 주변에 2차원 상자인 경계 상자를 그리는 과정이 포함됨
- 레이블링된 경계 상자가 있는 이미지를 통해 훈련된 컴퓨터 비전 모델은 상자 내의 픽셀이 지정된 객체를 해당한다는 것을 학습함



③ 의미 체계 세분화

- 의미 체계 세분화를 사용하면 모델이 학습해야 하는 레이블에 해당하는 정확한 이미지 부분을 레이블링할 수 있음
- 개별 픽셀이 레이블링 되기 때문에 고도로 정밀한 훈련 데이터가 생성



- 텍스트

- ✓ 텍스트 데이터에 대한 레이블링 워크플로우를 기본적으로 제공

① 텍스트 분류

- 텍스트 분류에는 사전 정의된 레이블 세트를 기준으로 텍스트 문자열을 분류하는 과정이 포함
- 텍스트를 각기 다른 레이블로 분류하는 작업은 주제(예: 제품 설명, 영화 리뷰), 감정 등을 식별하는 자연어 처리(NLP) 모델에 사용되는 경우가 많음

Hello Julia Log out

Instructions ×

[View full instructions](#)

[View tool guide](#)

Positive
Positive means some aspects of the text uncover a positive mood, such as praise, pleasure, recommendation or a favorable comparison.
Example: This view is amazing.

Negative
Negative means some aspects of the text uncover a negative mood, such as criticism, insults or a negative comparison.
Example: I dislike old cabin cruisers.

Neutral
Neutral means no emotions are implied.
Example: It's going to rain tomorrow.

Unsure
Select this option when you are not sure what sentiment the content is.
Source: BDU

What is the overall sentiment in the text?

My boyfriend and I went to watch The Guardian. At first I didn't want to watch it, but I loved the movie- It was definitely the best movie I have seen in sometime. They portrayed the USCG very well, it really showed me what they do and I think they should really be appreciated more. Not only did it teach but it was a really good movie. The movie shows what the really do and how hard the job is. I think being a USCG would be challenging and very scary. It was a great movie all around. I would suggest this movie for anyone to see. The ending broke my heart but I know why

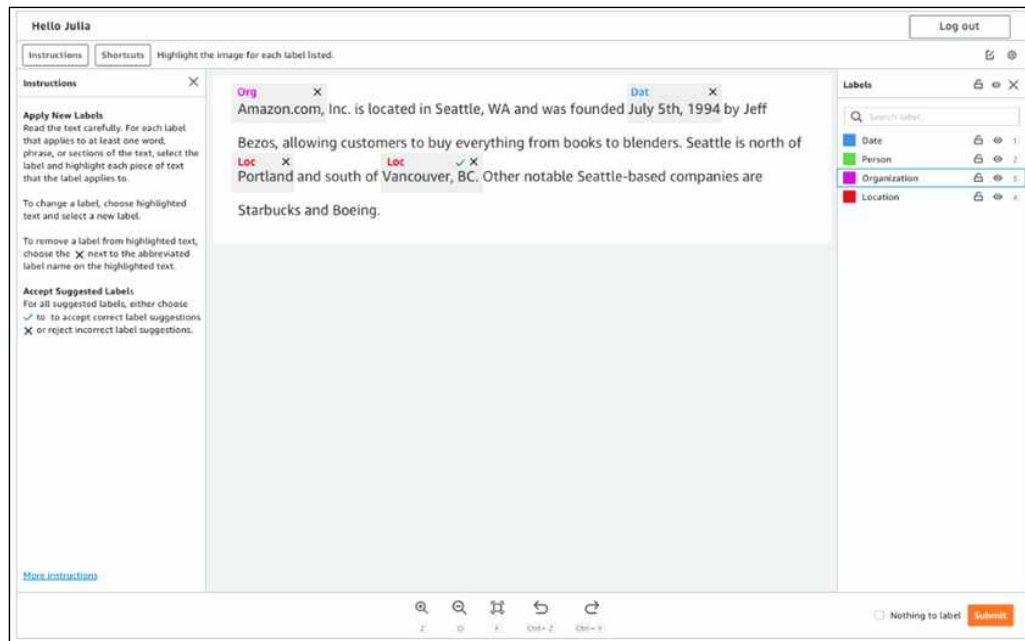
Select an option

<input checked="" type="radio"/> Positive	1
<input type="radio"/> Negative	2
<input type="radio"/> Neutral	5
<input type="radio"/> Unsure	4

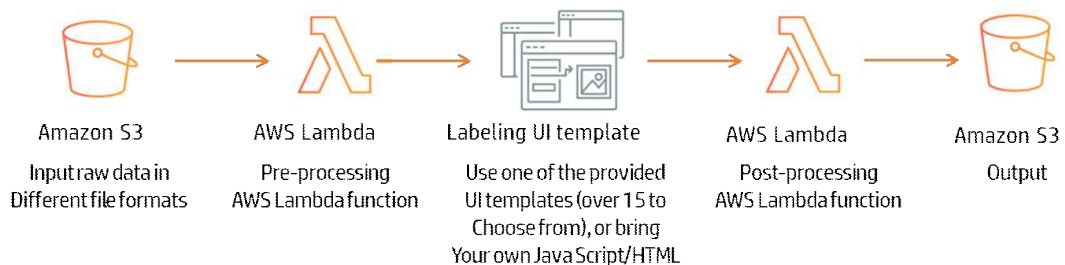
Submit

② 명명된 엔터티 인식

- 명명된 엔터티 인식(NER)에는 텍스트 데이터를 조사하여 명명된 엔터티라는 구문을 찾고 이러한 각 구문을 “사람”, “조직” 또는 “브랜드” 같은 레이블로 분류하는 과정이 포함



- 사용자 지정 워크플로우
 - ✓ Ground Truth에서는 자체 레이블링 워크플로우를 생성할 수 있음
 - ✓ 사용자 지정 워크플로우 구성 요소
- ① 작업자에게 레이블링 태스크를 완료하는 데 필요한 모든 지침과 도구를 제공하는 UI
 - ② AWS Lambda 함수에 캡슐화된 사전 처리 로직
 - ③ AWS Lambda 함수에 캡슐화된 사후 처리 로직



- 인력
 - ✓ 세이지메이커 Ground Truth에서는 여러 가지 옵션을 통해 인력이 데이터를 레이블링하도록 지원하는 옵션들이 있음
 - ✓ 옵션의 종류
 - ① 자체 직원
 - ② AWS Marketplace를 통해 제공되는 타사 데이터 레이블링 서비스 공급자
 - ③ Amazon Mechanical Turk를 통한 클라우드 소싱 인력

평가하기

1. 세이지메이커 Ground Truth의 작동 순서를 올바르게 나열하시오.

- ① 원시 데이터
- ② 작업자
- ③ 보조 레이블링
- ④ 레이블링

- 정답 : ① → ④ → ② → ③ 번

해설 : SageMaer Ground Truth는 원시데이터 → 레이블링 →작업자 → 보조 레이블링 동작을 수행 후 모델 훈련에 사용할 데이터 세트 준비가 완료 됩니다.

2. 세이지메이커 Ground Truth의 이점으로 옳지 않은 것을 고르시오.

- ① 사용 편의성
- ② 비용 절감
- ③ 인력 선택
- ④ 사용자 지정 워크플로우

- 정답 : ④번

해설 : 세이지메이커 Ground Truth의 이점으로는 데이터 레이블 정확도 개선, 사용 편의성, 비용 절감, 인력 선택이 있습니다. 사용자 지정 워크플로우는 세이지메이커 Ground Truth의 주요 기능입니다.

학습정리

1. 세이지메이커 Ground Truth 개요

- 기계학습을 위한 고도로 정확한 훈련 데이터 세트를 손쉽게 구축할 수 있게 해주는 완전 관리형 데이터 레이블링 시스템
- 작동 방식 :원시 데이터 → 레이블링 → 작업자 →보조레이블링

2. 세이지메이커 Ground Truth 이점

- 데이터 레이블 정확도 개선
- 사용 편의성
- 비용 절감
- 인력 선택

3. 세이지메이커 Ground Truth 주요 기능

- 3D 포인트 클라우드 : 객체 탐지, 객체 추적, 의미 체계 세분화
- 동영상 : 비디오 객체 감지, 비디오 객체 추적, 비디오 클립 분류
- 이미지 : 이미지 분류, 객체 탐지, 의미 체계 세분화
- 텍스트 : 텍스트 분류, 명명된 엔터티 인식
- 사용자 지정 워크플로우 : 작업자에게 레이블링 태스크를 완료하는데 필요한 모든 지침과 도구를 제공하는 UI, AWS Lambda 함수에 캡슐화 된 사전/사후 처리 로직
- 인력 : 자체직원, 타사 레이블링 서비스 공급자, 클라우드 소싱 인력

