

# AI 평가 지표

# 정확도 (Accuracy)

# 정확도 (Accuracy)

올바르게 예측된 데이터의 수

---

전체 데이터 수

**올바르게 암 발생 유무가 예측된 인원**

---

**암 건강검진 총 인원**

**90,000**



**100,000**

**90,000**



**100,000**



**정확도 : 90%**

**이런 엄청난 AI 모델 사실래요?**  
**(다른 회사 모델은 70%대임)**

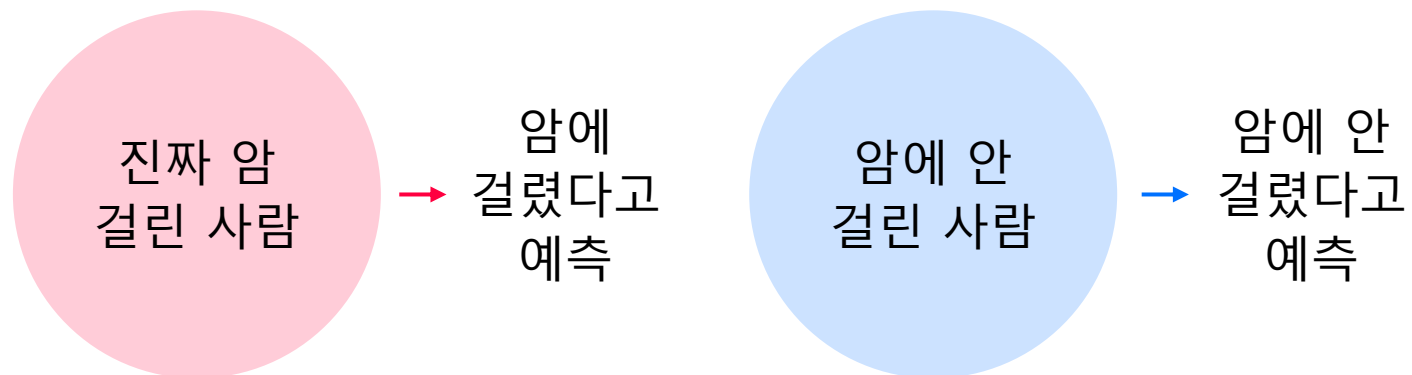
10/10 ...



**올바르게 암 발생 유무가 예측된 인원**

---

**암 건강검진 총 인원**



**올바르게 암 발생 유무가 예측된 인원**

---

**암 건강검진 총 인원**

**90,000**



**100,000**

맞추지 못한 10,000명은?

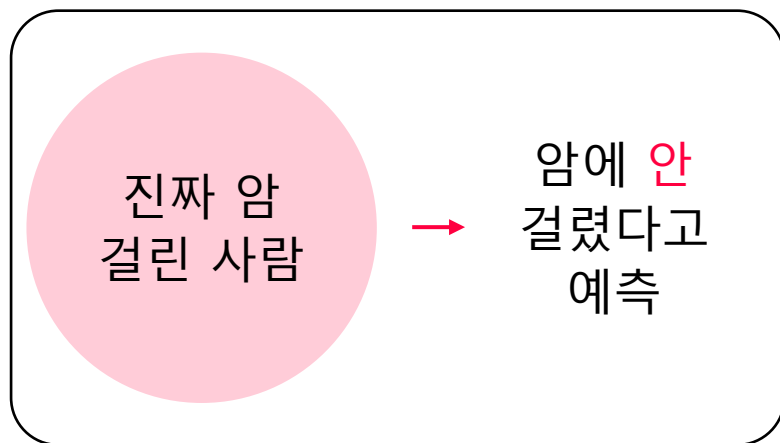
90,000



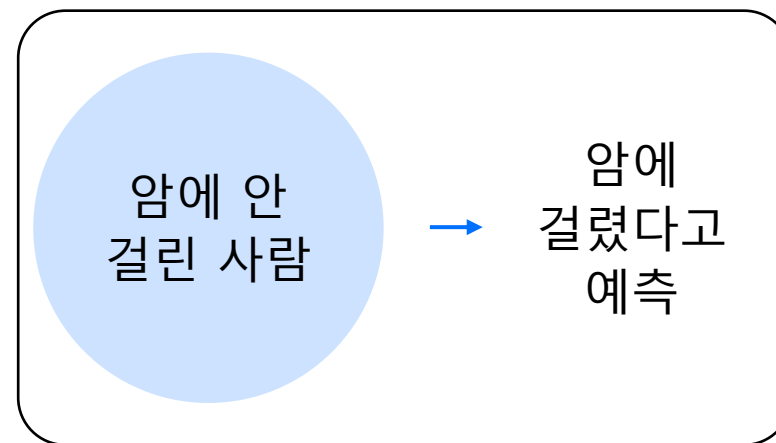
---

100,000

## 오답인 10,000명의 데이터 확인

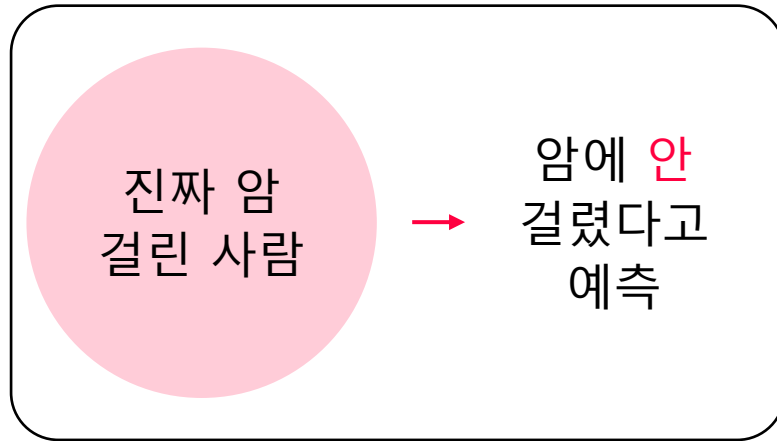


9,000

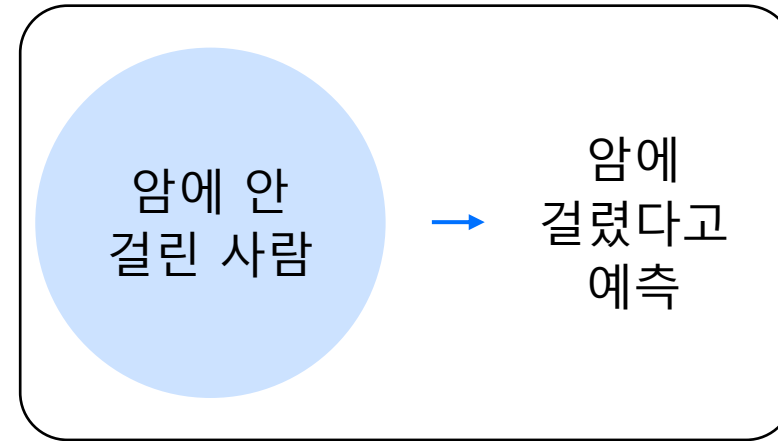


1,000

## 오답인 10,000명의 데이터 확인



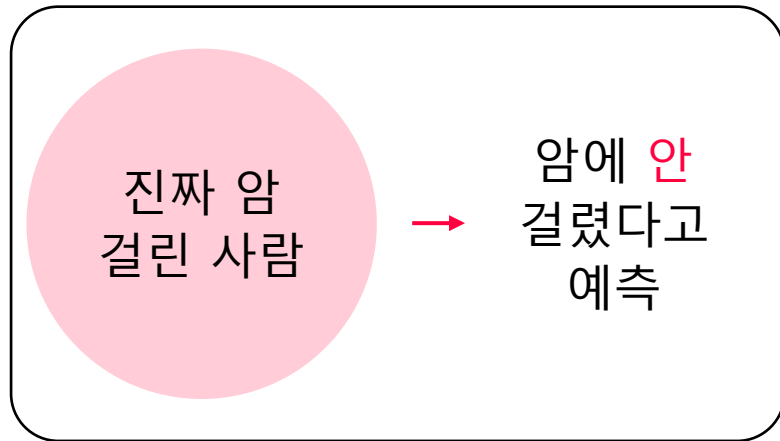
9,000



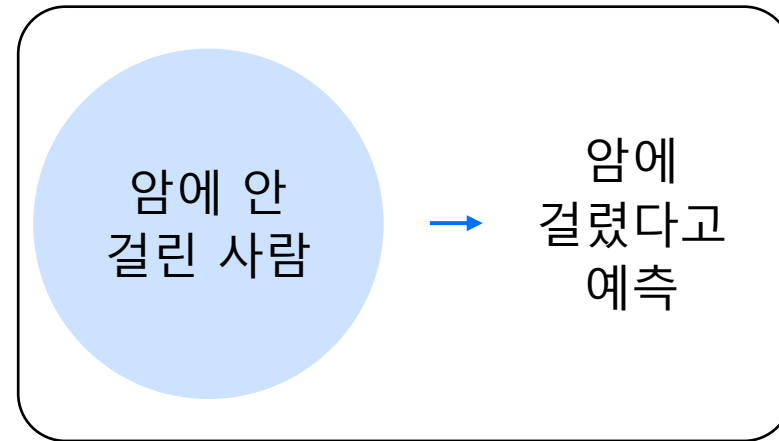
1,000

대부분 암이 아니라고 판단

## 오답인 10,000명의 데이터 확인



9,000

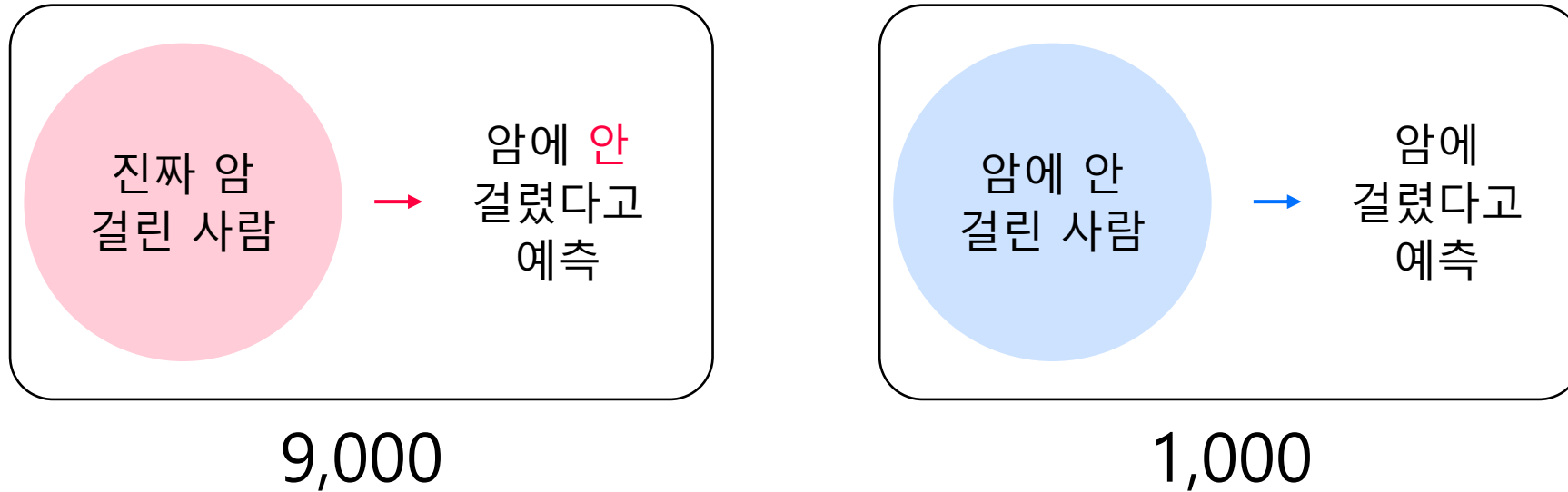


1,000

대부분 암이 아니라고 판단

✓ 일반적으로 암에 걸리지 않았을 확률이 높고,

## 오답인 10,000명의 데이터 확인

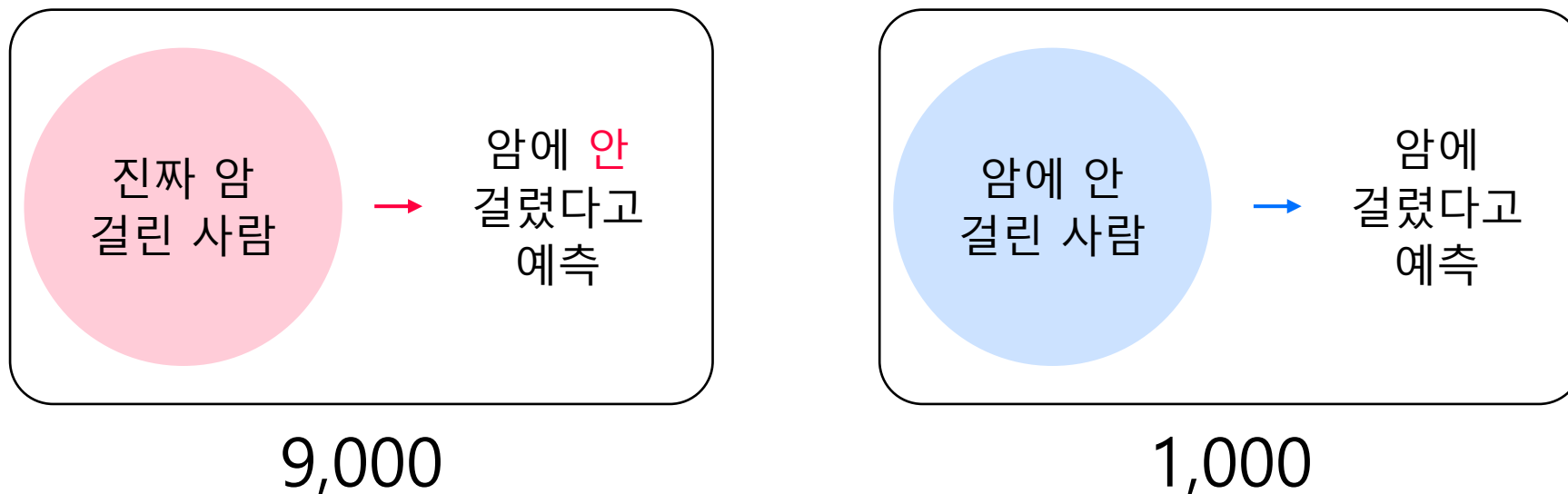


대부분 암이 아니라고 판단

- ✓ 일반적으로 암에 걸리지 않았을 확률이 높고,
- ✓ 학습 데이터셋도 불균형했음



## 오답인 10,000명의 데이터 확인



대부분 암이 아니라고 판단

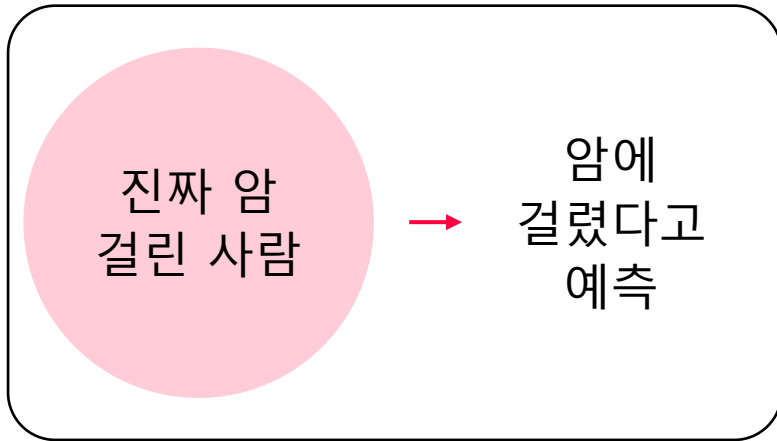
- ✓ 일반적으로 암에 걸리지 않았을 확률이 높고,
  - ✓ 학습 데이터셋도 불균형했음
- ✓ 이 AI는 사실 대부분 암이 아니라고 판단하는 AI였던 것

**90,000**

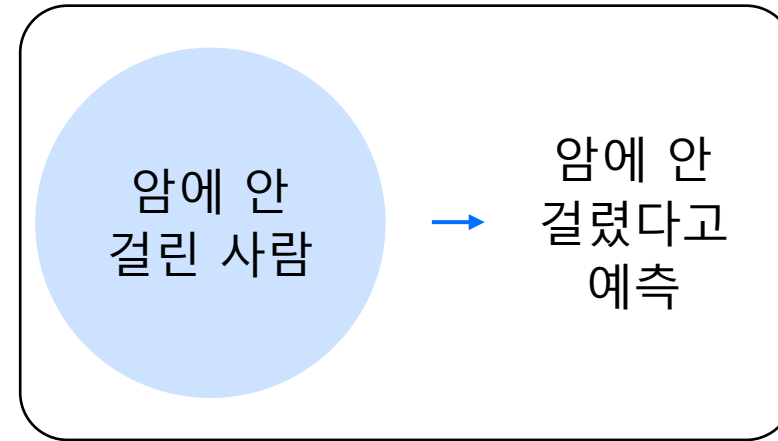
---

**100,000**

## 정답인 90,000명의 데이터 확인



5,000



85,000

**이런 엄청난 AI 모델 사실래요?**  
**(다른 회사 모델은 70%대임)**

# 이런 엄청난 AI 모델 사실래요?



사면 절대 안됩니다  
그냥 암에 안걸렸다고 때려 맞추는 수준

**즉, 정확도는 한계가 있음**

# 재현율(Recall), 정밀도(Precision)

# 재현율(Recall), 정밀도(Precision)

		실제 정답	
		Positive	Negative
예측 결과	Positive	True Positive	False Positive
	Negative	False Negative	True Negative



# 재현율(Recall), 정밀도(Precision)

		실제 정답	
		True	False
예측 결과	True	진짜를 진짜라고 예측	False Positive
	False	False Negative	가짜를 가짜라고 예측

# 재현율(Recall), 정밀도(Precision)

		실제 정답	
		True	False
예측 결과	True	진짜를 진짜라고 예측	False Positive
	False	False Negative	가짜를 가짜라고 예측

# 재현율(Recall), 정밀도(Precision)

		실제 정답	
		True	False
예측 결과	True	진짜를 진짜라고 예측	진짜를 가짜라고 예측
	False	가짜를 진짜라고 예측	가짜를 가짜라고 예측

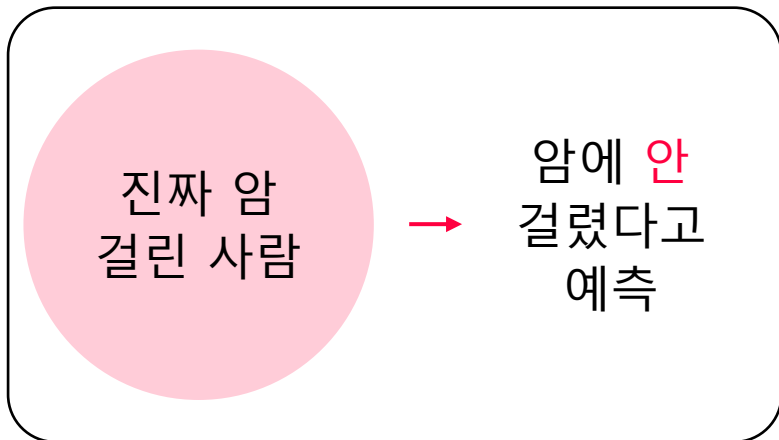
# 재현율(Recall), 정밀도(Precision)

		실제 정답	
		True	False
예측 결과	True	암환자를 암환자라고 예측(TP)	암환자가 아닌 사람을 암환자라고 예측(FP)
	False	암환자를 암환자가 아니라고 예측(FN)	암환자가 아닌 사람을 암환자가 아니라고 예측 (TN)

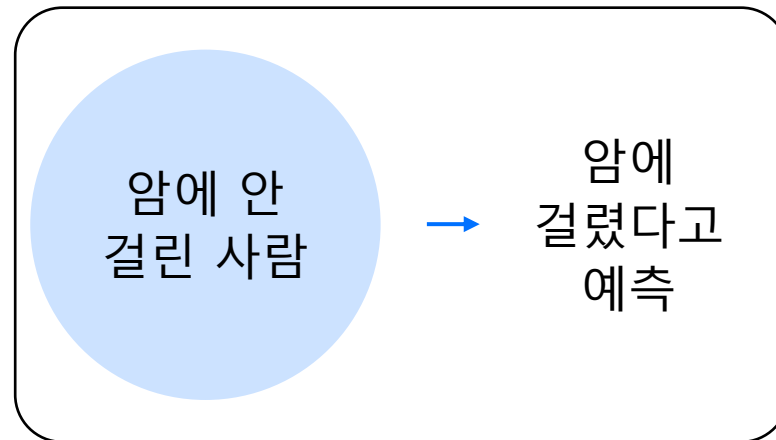
# 재현율(Recall), 정밀도(Precision)

예측 결과		실제 정답	
		True	False
		True	False
예측 결과	True	5,000	1,000
	False	9,000	85,000

## 오답인 10,000명의 데이터 확인



9,000(FN)



1,000(FP)

# 재현율(Recall), 정밀도(Precision)

		실제 정답	
		True	False
예측 결과	True	5,000	1,000
	False	9,000	85,000

# 정답인 90,000명의 데이터 확인

5,000(TP)

85,000(TN)



# 재현율(Recall), 정밀도(Precision)

		실제 정답	
		True	False
예측 결과	True	5,000	1,000
	False	9,000	85,000

# 재현율(Recall), 정밀도(Precision)

예측 결과		실제 정답	
		True	False
		True	False
True	5,000	1,000	
False	9,000	85,000	

# 재현율(Recall)

실제로 True인 것 중에서, 모델이 True라고 인식한 데이터의 비율

		실제 정답	
		True	False
예측 결과	True	5,000(TP)	1,000(FP)
	False	9,000(FN)	85,000(TN)

$$\frac{\text{TruePositives}}{\text{True Positives} + \text{False Negatives}}$$

# 재현율(Recall)

실제로 True인 것 중에서, 모델이 True라고 인식한 데이터의 비율

		실제 정답	
		True	False
예측 결과	True	5,000(TP)	1,000(FP)
	False	9,000(FN)	85,000(TN)

$$\frac{5,000}{5,000 + 9,000}$$

recall: 0.3571

# 재현율(Recall)

실제로 True인 것 중에서, 모델이 True라고 인식한 데이터의 비율

		실제 정답		
		True	False	
예측 결과	True	5,000	1,000	<div>5,000</div> <hr/> <div>5,000 + 9,000</div>
	False	9,000	85,000	

recall: 0.3571

암에 걸린 사람이 14,000명인데, 9,000명을 놓치고 있음

# 정밀도(Precision)

True로 예측한 것 중, 실제로 True인 데이터의 비율

		실제 정답	
		True	False
예측 결과	True	5,000(TP)	1,000(FP)
	False	9,000(FN)	85,000(TN)

$$\frac{\text{TruePositives}}{\text{True Positives} + \text{False Positive}}$$

# 정밀도(Precision)

True로 예측한 것 중, 실제로 True인 데이터의 비율

		실제 정답	
		True	False
예측 결과	True	5,000(TP)	1,000(FP)
	False	9,000(FN)	85,000(TN)

$$\frac{5,000}{5,000 + 1,000}$$

Precision: 0.8333

# 정밀도(Precision)

True로 예측한 것 중, 실제로 True인 데이터의 비율

		실제 정답		
		True	False	
예측 결과	True	5,000	1,000	$\frac{5,000}{5,000 + 1,000}$
	False	9,000	85,000	

Precision: 0.8333

1,000명의 사람이 암이 아닌데, 암 진단을 받아 2차 검사를 받음



**우리가 검사 받는다고 생각해봅시다.**

**암인줄 알았는데, 실제론 암이 아니다**

**암이 아닌줄 알았는데, 실제로 암이다**

**우리가 검사 받는다고 생각해봅시다.**

**암인줄 알았는데, 실제론 암이 아니다**

**암이 아닌줄 알았는데, 실제로 암이다**

# 재현율(Recall)

실제로 True인 것 중에서, 모델이 True라고 인식한 데이터의 비율

		실제 정답		
		True	False	
예측 결과	True	5,000	1,000	<div>5,000</div> <hr/> <div>5,000 + 9,000</div>
	False	9,000	85,000	

recall: 0.3571

암에 걸린 사람이 14,000명인데, 9,000명을 놓치고 있음

**스팸 메일인줄 알았는데, 실제론 스팸 메일이 아니다.**

**스팸 메일이 아닌 줄 알았는데 ,실제로 스팸 메일이다.**

**스팸 메일인줄 알았는데, 실제론 스팸 메일이 아니다.**

중요 메일을 스팸 메일로 구분함 -> 와....

**스팸 메일이 아닌 줄 알았는데, 실제로 스팸 메일이다.**

스팸 메일을 중요 메일로 구분함 -> 귀찮음

# 정밀도(Precision)

True로 예측한 것 중, 실제로 True인 데이터의 비율

		실제 정답		
		True	False	
예측 결과	True	5,000	1,000	$\frac{5,000}{5,000 + 1,000}$
	False	9,000	85,000	

Precision: 0.8333

1,000명의 사람이 암이 아닌데, 암 진단을 받아 2차 검사를 받음

재현율, 정밀도 중 어떤 것이 중요한지는  
테스크에 따라 다르다

# 재현율과 정밀도는 Trade-off 관계

재현율이 극도로 높아지면, 정밀도가 매우 낮아짐

정밀도가 극도로 높아지면, 재현율이 매우 낮아짐



이를 해결하기 위한 F1 Score

## 이를 해결하기 위한 F1 Score

$$2* \frac{\text{재현율} * \text{정밀도}}{\text{재현율} + \text{정밀도}}$$

## 이를 해결하기 위한 F1 Score

재현율: 0.82

정밀도: 0.5

▶ F1 Score: 0.62

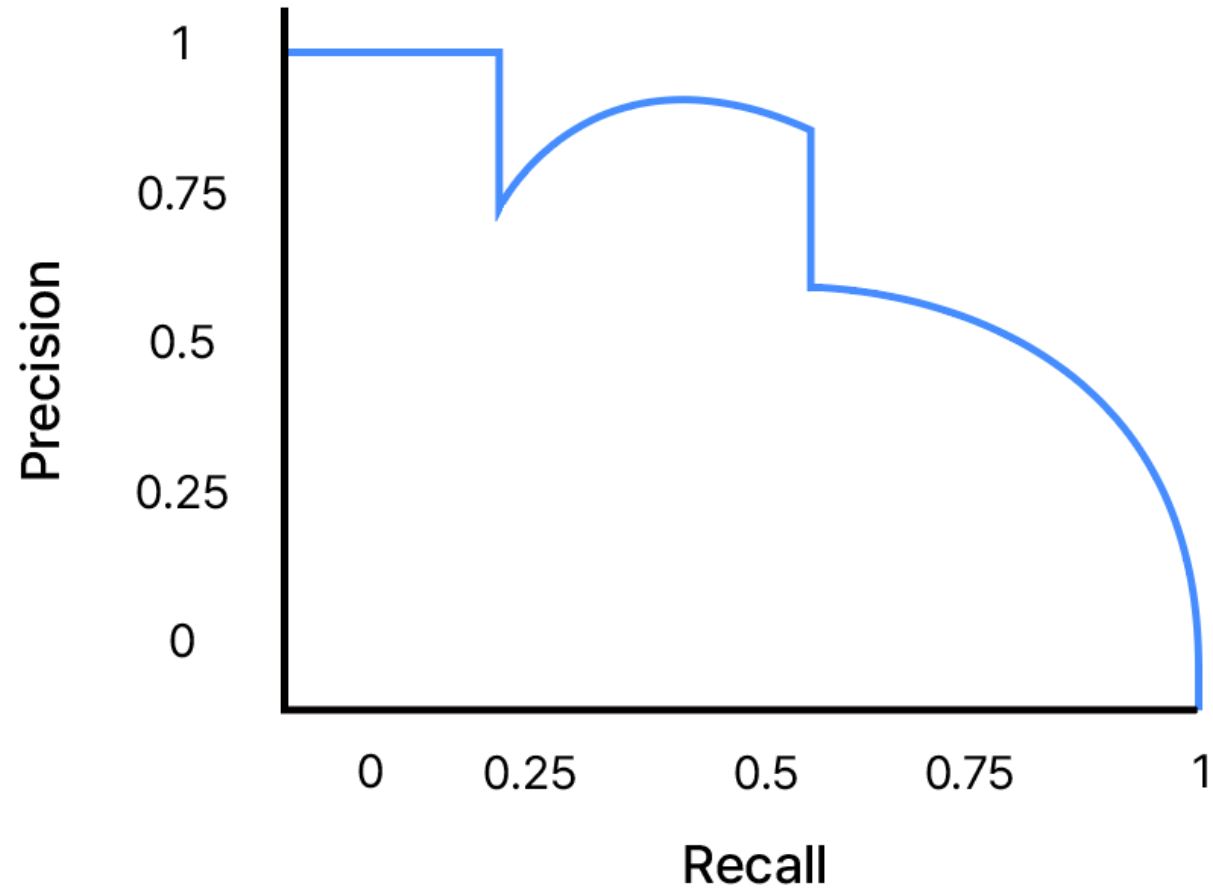
재현율: 0.65

정밀도: 0.65

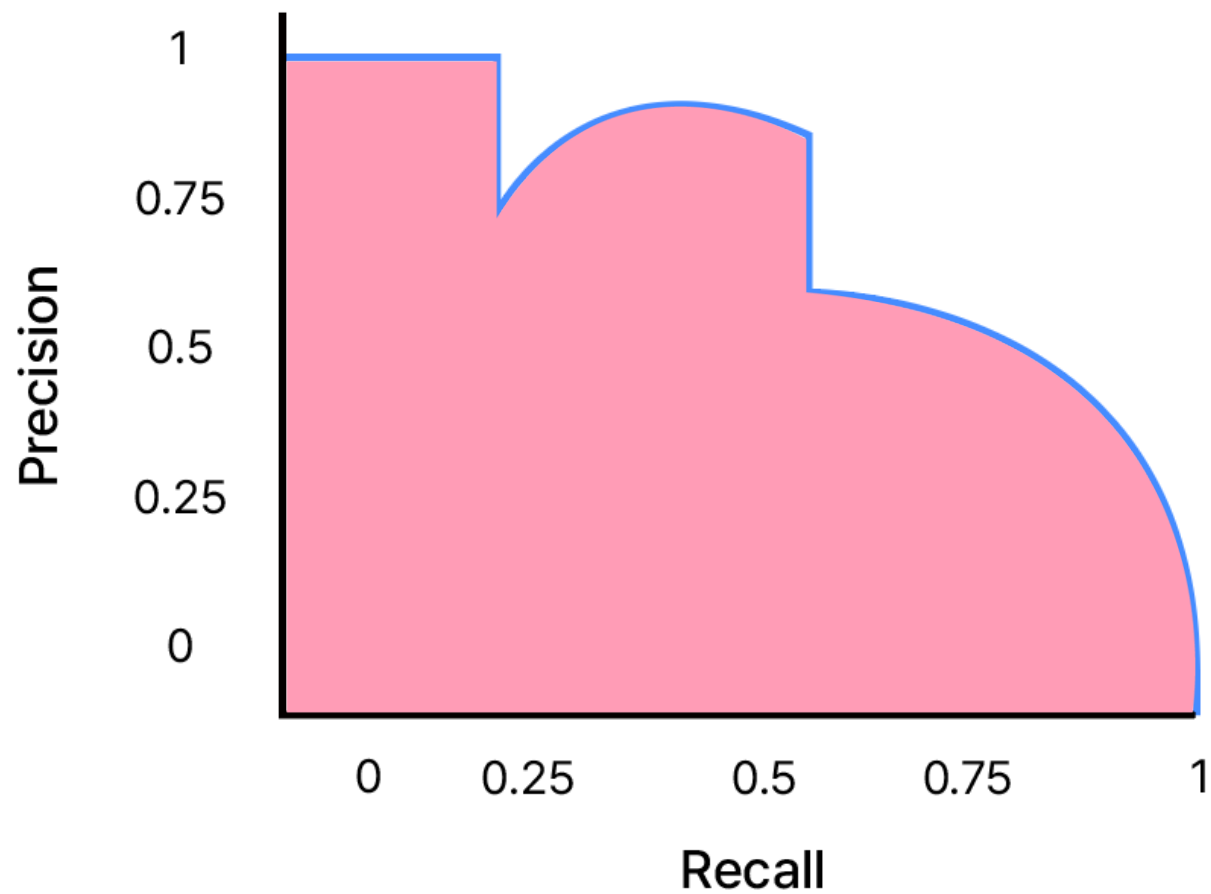
▶ F1 Score: 0.65

# Precision-Recall Curve

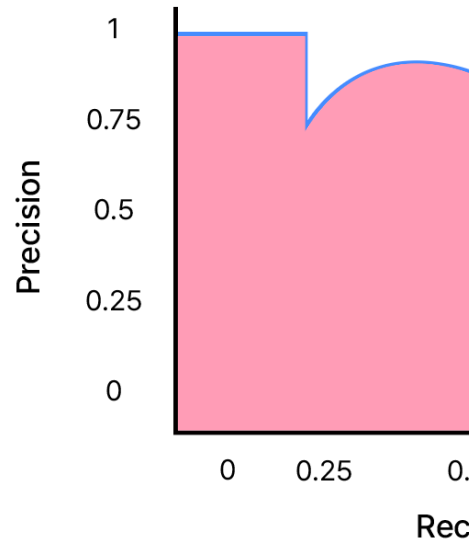
# Precision-Recall Curve



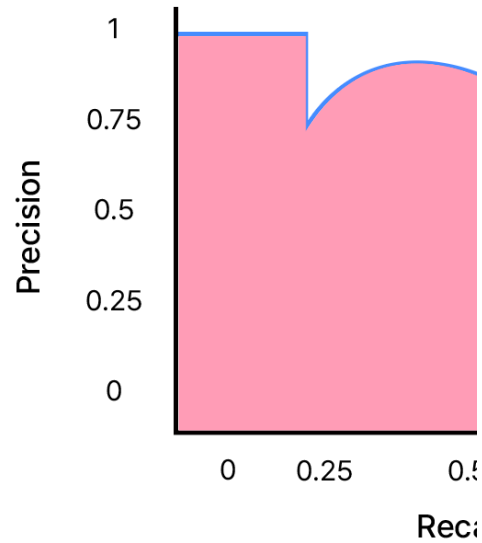
# AP(Average Precision)



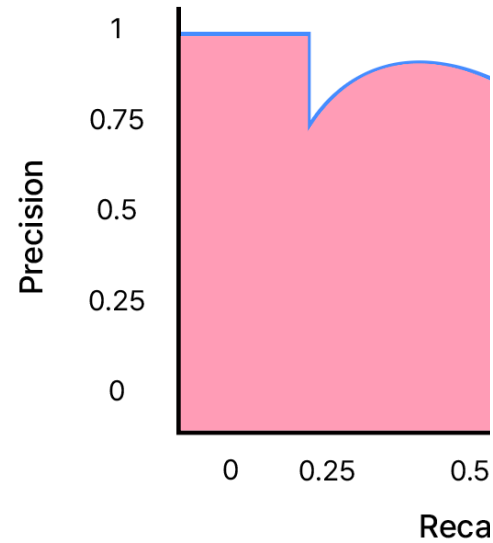
# MAP(Mean Average Precision)



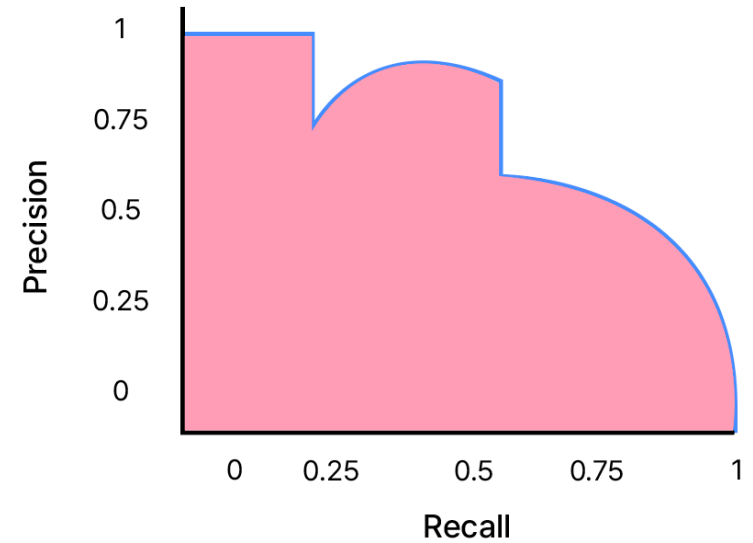
강아지



고양이



새



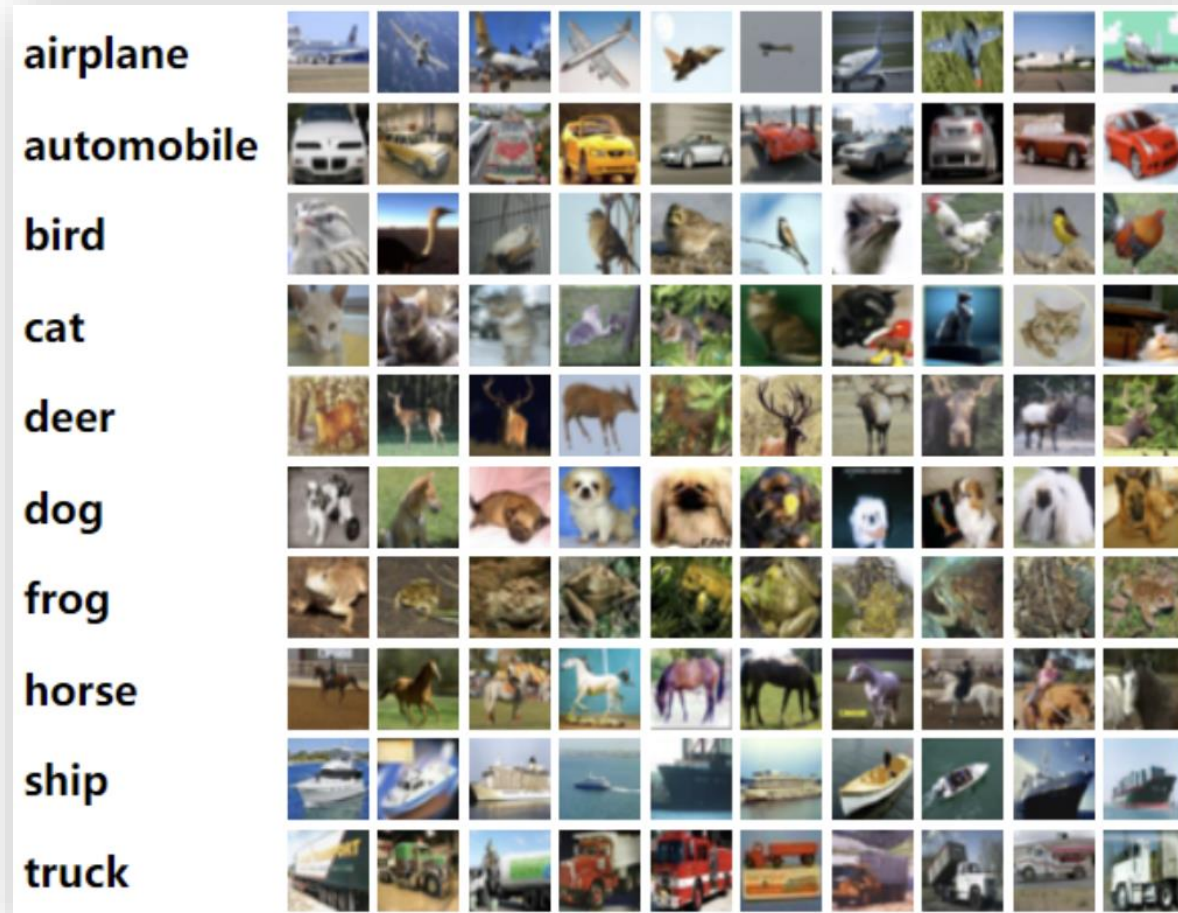
물고기

각 클래스들간의 AP 평균

# TOP-5 Accuracy



# TOP-5 Accuracy



총 10개의 클래스가 있는 CIFAR-10 데이터셋

# TOP-5 Accuracy



# TOP-5 Accuracy



예측해봐

비행기	자동차	새	고양이	사슴	강아지	개구리	말	배	트럭
0.15	0.06	0.07	0	0	0.04	0.05	0.01	0.6	0.02



## TOP-5 Accuracy



확률값이 높게 나온 5개 중,  
진짜 정답이 있으면 답으로 인정

비행기	자동차	새	고양이	사슴	강아지	개구리	말	배	트럭
0.15	0.06	0.07	0	0	0.04	0.05	0.01	0.6	0.02

# TOP-5 Accuracy



7% 지만 Top-5 확률에 들었으므로,  
**정답으로 인정**

비행기	자동차	새	고양이	사슴	강아지	개구리	말	배	트럭
0.15	0.06	0.07	0	0	0.04	0.05	0.01	0.6	0.02

# TOP-1 Accuracy



확률값이 높게 나온 1개 중,  
진짜 정답이 있으면 답으로 인정

비행기	자동차	새	고양이	사슴	강아지	개구리	말	배	트럭
0.15	0.06	0.07	0	0	0.04	0.05	0.01	0.6	0.02

# TOP-1 Accuracy



Top-1에는 들지 못했으므로 ,  
**정답이 아님**

비행기	자동차	새	고양이	사슴	강아지	개구리	말	배	트럭
0.15	0.06	0.07	0	0	0.04	0.05	0.01	0.6	0.02



Q&A