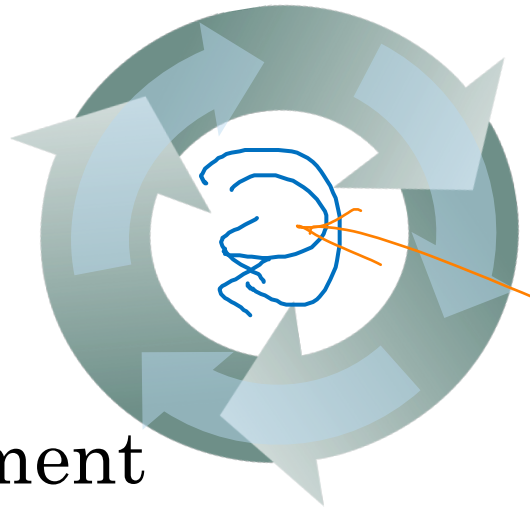
deeplearning.ai

Setting up your goal

Single number evaluation metric

어떤 개선 작업을 시도하건 간에 단일한 metric이 있으면 내가 새로 한 작업이 모델에 좋은 영향을 미치는지 아닌지 확실하게 판단할 수 있다.

Using a single number evaluation metric



Of examples recognized as cat, what % actually are cats?
examples recognized as cat
actually are cats

What % of actual cats are correctly recognized

Classifier	Precision	Recall	F1 score
이전 모델 A	95%	90%	92.4%
개선된 모델 B	98%	85%	91.0%

이 둘 중에 뭐가 좋은지 고민하는건 선택할 모델이 많을 때는 고된 일이 된다. 그냥 F1 score 같은 단일 metric을 써라.

F1 score = "Average" of P and R.

$\left(\frac{2}{\frac{1}{P}+\frac{1}{R}}\right.$. "Harmonic mean")

Dev set + Single number evaluation metric

real Speed up iterating

Andrew Ng

Another example

metric이 4개나 된다면 모델이 많을 때 뭘 선택할지 결정하기 쉽지 않다.
이럴 땐 average performance를 추가로 관리하면 된다.

Algorithm	US	China	India	Other	Average
A	3%	7%	5%	9%	6%
B	5%	6%	5%	10%	6.5%
C	2%	3%	4%	5%	3.5%
D	5%	8%	7%	2%	5.25%
E	4%	5%	2%	4%	3.75%
F	7%	11%	8%	12%	9.5%