



deeplearning.ai

Setting up  
your goal

---

When to change  
dev/test sets and  
metrics

# Cat dataset examples

Metric + Dev : Prefer A  
You/users : Prefer B.

evaluation metric이 진짜로 어떤 알고리즘이 더 좋은지 정확한 rank를 주지 못한다면 새로운 evaluation metric을 써야 할 때인거다.

→ Metric: classification error

Algorithm A: 3% error → Pornographic

✓ Algorithm B: 5% error

Error:  $\frac{1}{\sum_i \omega^{(i)}} - \frac{1}{m_{dev}}$

$$\sum_{i=1}^{m_{dev}} \underline{\omega^{(i)}} \downarrow \left\{ \frac{y_{pred}^{(i)} + y^{(i)}}{\text{predicted value (0/1)}} \right\}$$

$\rightarrow \omega^{(i)} = \begin{cases} 1 & \text{if } x^{(i)} \rightarrow \text{non-porn} \\ 10 & \text{if } x^{(i)} \rightarrow \text{porn} \end{cases}$

포르노 사진에 weight를 크게 줘서 학습에 페널티를 준다.

# Orthogonalization for cat pictures: anti-porn

- 1. So far we've only discussed how to define a metric to evaluate classifiers. ← Place target  completely separate knobs 
- 2. Worry separately about how to do well on this metric. 

Aim (shoot at target)  
aim / shoot

$$\rightarrow J = \frac{1}{m} \sum_{i=1}^m \ell(\hat{y}^{(i)}, y^{(i)})$$

2번 관점에서 알고리즘이 최적화할 cost function을 정하고  
1번 관점에서 이것을 바꾸는 것을 완전히 별개의 과정으로 수행하자.



# Another example

Algorithm A: 3% error

✓ Algorithm B: 5% error ←

→ Dev/test



→ User images



If doing well on your metric + dev/test set does not correspond to doing well on your application, change your metric and/or dev/test set.

고품질 이미지의 dev/test set으로 evaluation을 수행했지만 사실 유저가 관심있는건 저런 저품질 이미지에서 성능이 잘 나오는가 이다.  
이럴땐 알고리즘이 실제로 더 잘해야 하는 것을 capture할 수 있도록 metric을 바꾸거나 dev/test set을 바꿔야 한다.