



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.

→ Metric: classification error

Algorithm A: 3% error

→ pornographic

✓ Algorithm B: 5% error

Error:  $\frac{1}{\sum_i w^{(i)}} \times \frac{1}{m_{dev}} \sum_{i=1}^{m_{dev}} w^{(i)} \mathbb{I}\{y_{pred}^{(i)} \neq y^{(i)}\}$

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

$\mathbb{I}\{y_{pred}^{(i)} \neq y^{(i)}\}$   
predicted value (0/1)

# Orthogonalization for cat pictures: anti-porn

- 1. So far we've only discussed how to define a metric to evaluate classifiers. ← Place target ↗
- 2. Worry separately about how to do well on this metric. ↗
- ↖ Aim (shoot at target)

$$\rightarrow J = \frac{1}{\sum w^{(i)}} \sum_{i=1}^m w^{(i)} \mathcal{L}(\hat{y}^{(i)}, y^{(i)})$$



# 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.