



deeplearning.ai

Setting up  
your goal

---

When to change  
dev/test sets and  
metrics

# Cat dataset examples

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

→ Metric: classification error

Algorithm A: 3% error

→ pornographic

✓ Algorithm B: 5% error

$$\left\{ \begin{array}{l} \text{Error: } \frac{1}{\sum_i w^{(i)}} \quad \frac{1}{m_{\text{dev}}} \quad \sum_{i=1}^{m_{\text{dev}}} w^{(i)} \quad \frac{1}{2} \{ y_{\text{pred}}^{(i)} \neq y^{(i)} \} \\ \rightarrow w^{(i)} = \begin{cases} 1 & \text{if } x^{(i)} \text{ is non-porn} \\ 10 & \text{if } x^{(i)} \text{ is porn} \end{cases} \end{array} \right.$$

$\underbrace{\quad}_{\text{predicted value (0/1)}}$

# Orthogonalization for cat pictures: anti-porn

- 1. So far we've only discussed how to define a metric to evaluate classifiers. ← Place target 20
- 2. Worry separately about how to do well on this metric. 20
- ↑ 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

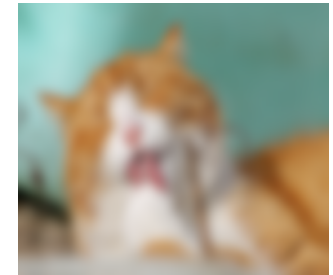
Algorithm A: 3% error

✓ 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.