



deeplearning.ai

Setting up
your goal

When to change
dev/test sets and
metrics

Cat dataset examples

When your evaluation metric is no longer correctly rank ordering preferences between algorithms, then that's a sign that you should change your evaluation metric or perhaps your development set or test set.

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

→ Metric: classification error

Algorithm A: 3% error

→ pornographic

✓ Algorithm B: 5% error

add the weight term here

$$\left\{ \begin{array}{l} \text{Error: } \frac{1}{\sum_i w^{(i)}} \quad \frac{1}{m_{\text{dev}}} \quad \sum_{i=1}^{m_{\text{dev}}} \frac{w^{(i)}}{1} \quad \mathbb{I} \{ \underbrace{y_{\text{pred}}^{(i)} \neq y^{(i)}}_{\substack{\text{predicted value (0/1)}}} \} \\ \rightarrow w^{(i)} = \begin{cases} 1 & \text{if } x^{(i)} \text{ is non-porn} \\ 10 & \text{if } x^{(i)} \text{ is porn} \end{cases} \end{array} \right.$$

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)

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

incorporate these weights



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.