

## Setting up your goal

When to change dev/test sets and metrics

## Cat dataset examples

Motoric + Der: Prefor A. Youlusons: Prefor B.

→ Metric: classification error

Algorithm A: 3% error

bornodrobyic

/ Algorithm B: 5% error

## Orthogonalization for cat pictures: anti-porn

→ 1. So far we've only discussed how to define a metric to evaluate classifiers. — Place toget to

 $\rightarrow$  2. Worry separately about how to do well on this metric.





## Another example

Algorithm A: 3% error

✓ Algorithm B: 5% error ←





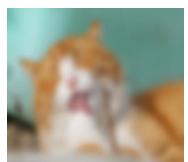




→ User images







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.