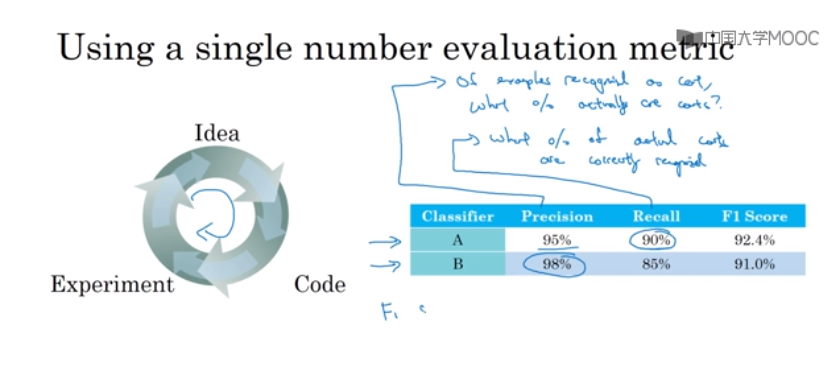


Orthogonalization: One knob tuning one thing.

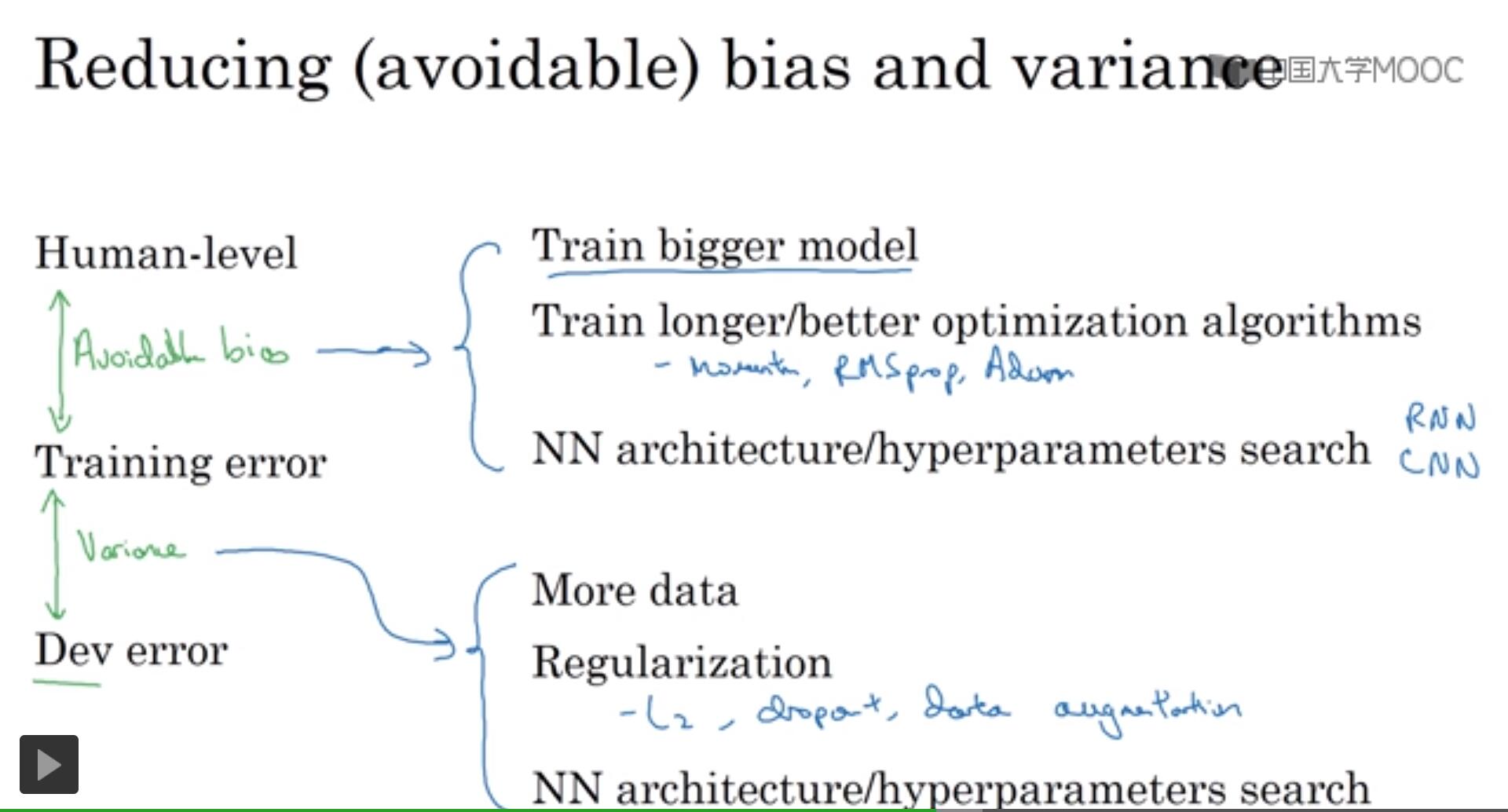
Example: early stopping is not good knob to use, because it change the performance both on training set and dev set.



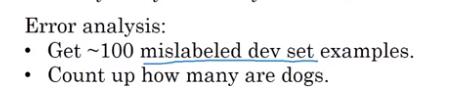
Precision：the probability of one picture is a cat.

Recall: What percentage of actual cats.

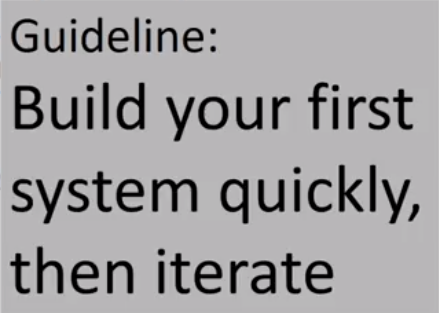
F1 Score: average of precision and recall.

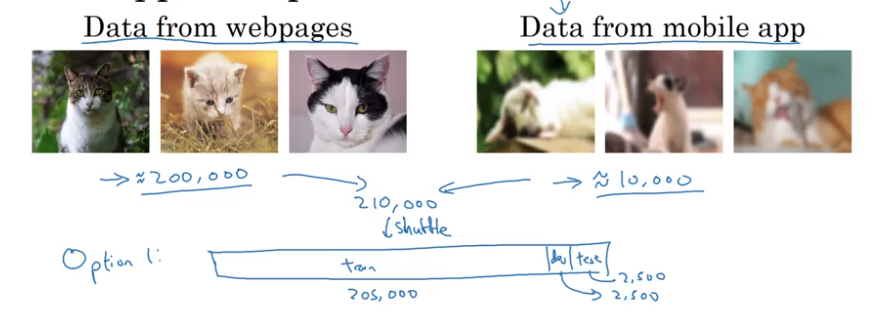


Error analysis



Dev/test set should come same distribution.





Option1:

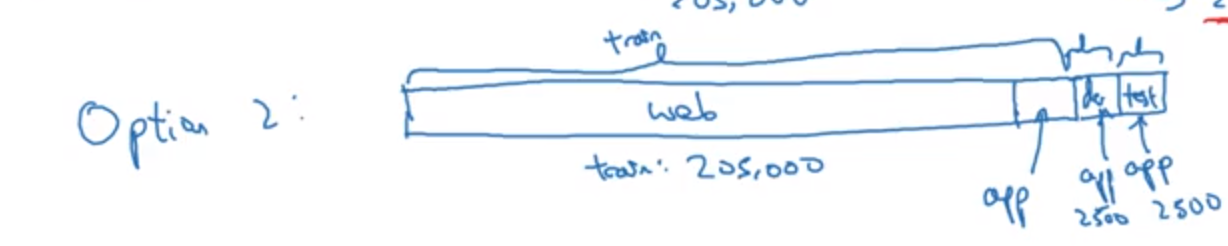
Advantage : training, dev, test sets will all come same distribution

Disvantage : dev set, lot of them come from web page.

Setting up dev set is telling your team where to aim the target.

Not good.

Option2:



Advantage: Aiming the target what you want to be.

Disadvantage: Training is different from test/dev distribution.

