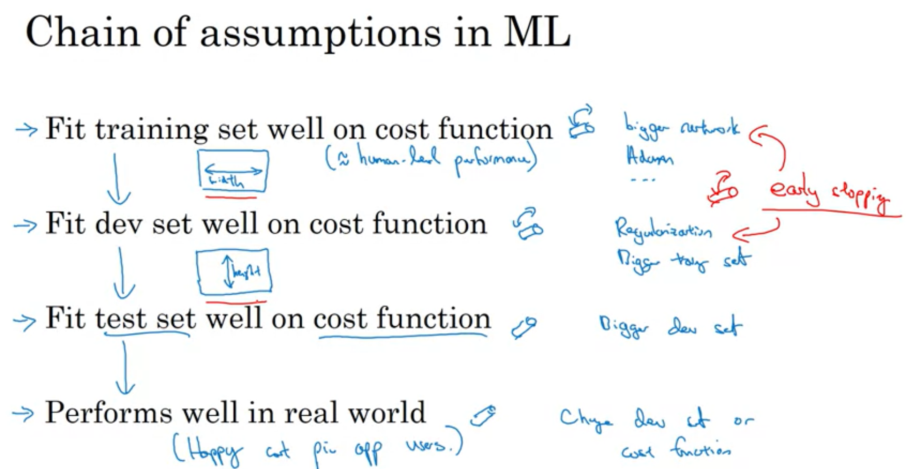
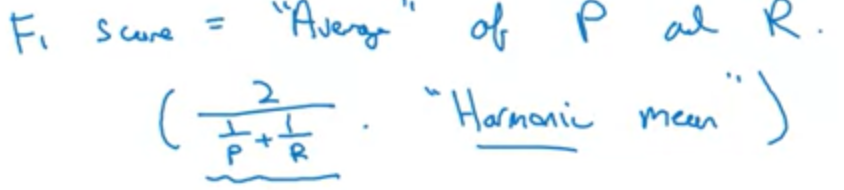
1. Orthogonalization：正交化

One knob controls one thing

try not to use early stop because it is changing two things together.

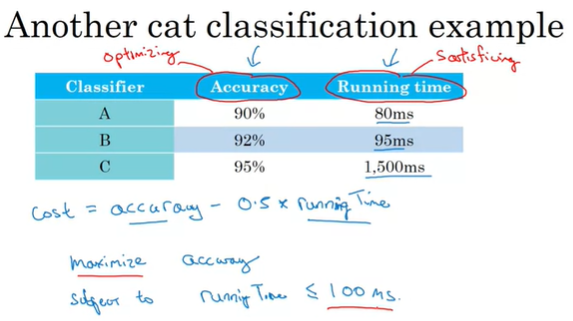
2. combine precision rate and recall rate: F1 score



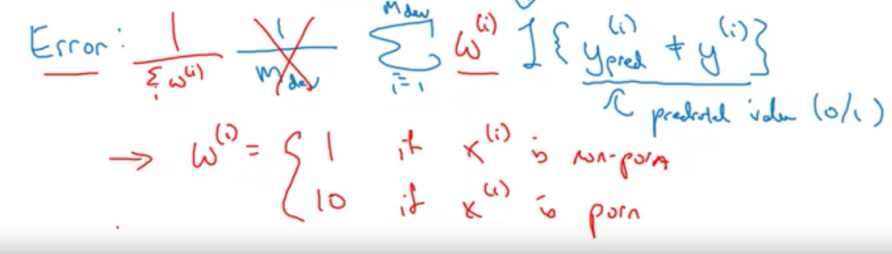
3. 在比较不同算法或者不同parameter时都尽量找到一个metric评价全局

4. 很难找到一个metric时，我们用optimize metric 和 satisfying metric

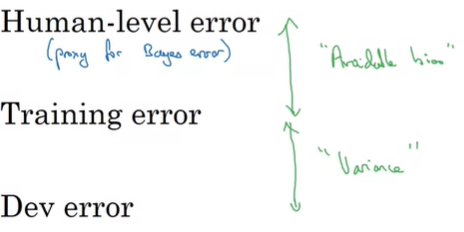
确定1个优化目标，其他的都是条件限制



5. 权重调整

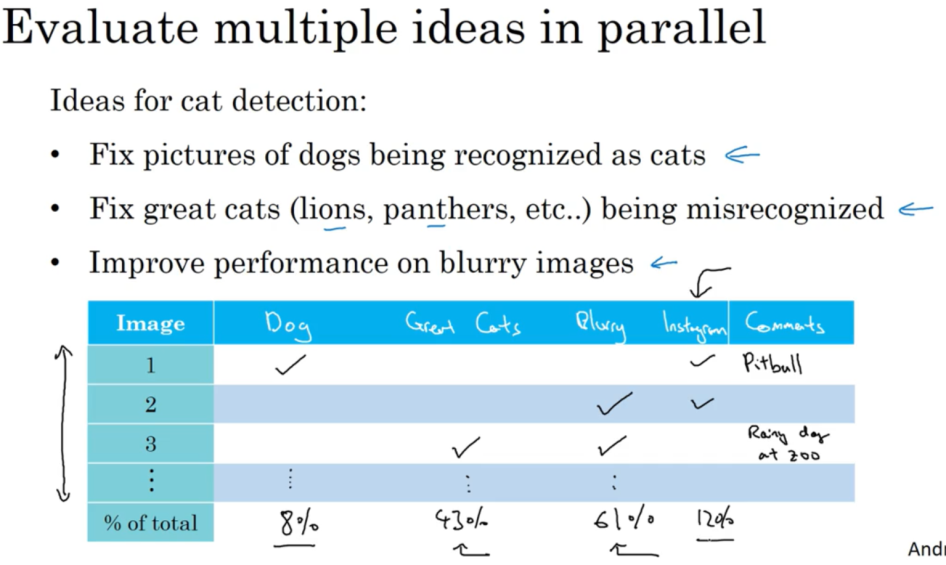
把不想分错的集权重调大

6. avoidable bias：使ML算法接近bayes error（proxy：人工错误率），可以避免过拟合。

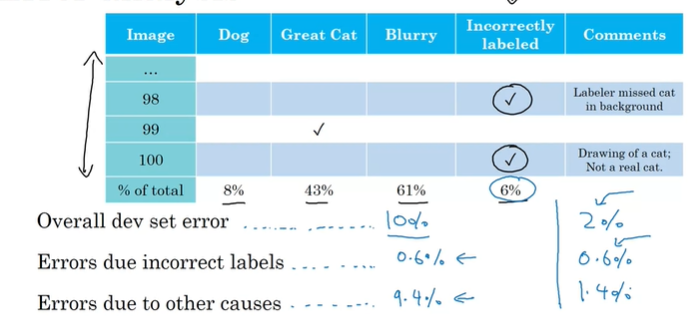


7. error analysis

人工看哪些错的多，增加在这些上面的研究

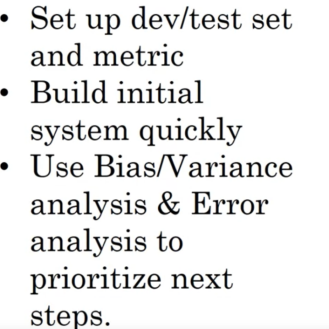


Mislabeled data：worth to fix or not？ Test/dev

第一列可以不考虑修正，第二列要考虑修正；training set 的randomly wrongly labeled可以不管

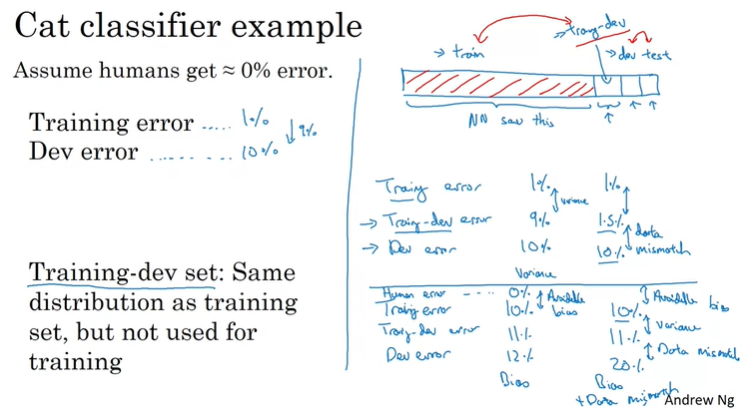
此处只统计了分错的里面的，但实际上分正确的理论上也要考虑，但实际中因为分对的太多，所以不好纠正

8.

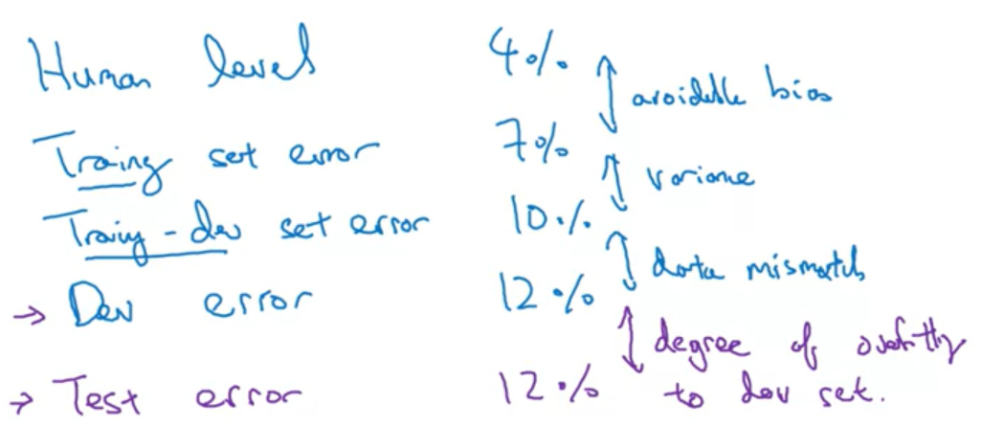
metric is important, it shows the target

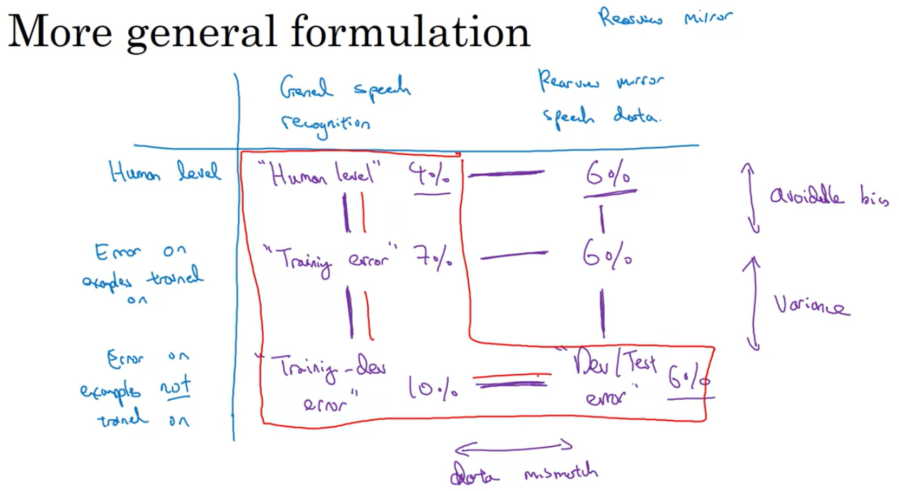
9. 如果数据来源的分布不一样，把最想要分好的分在test/dev set，training set只加一部分非常想分好的数据。

这样导致的问题是：当训练完一个模型在dev上运行时，可能出现的错误会很多因为train和dev还有test来自不同的分布，但同时还有可能是variance问题。所以对此我们引入下面的新概念：training-traindev-dev-test, 四个训练集。其中traindev是从原training中分出来的



Traindev error 成为判断有没有data-mismatch问题的关键；其与training error相近说明就是data mismatch问题；如果差的远就是variance问题



这是一个完整的info table

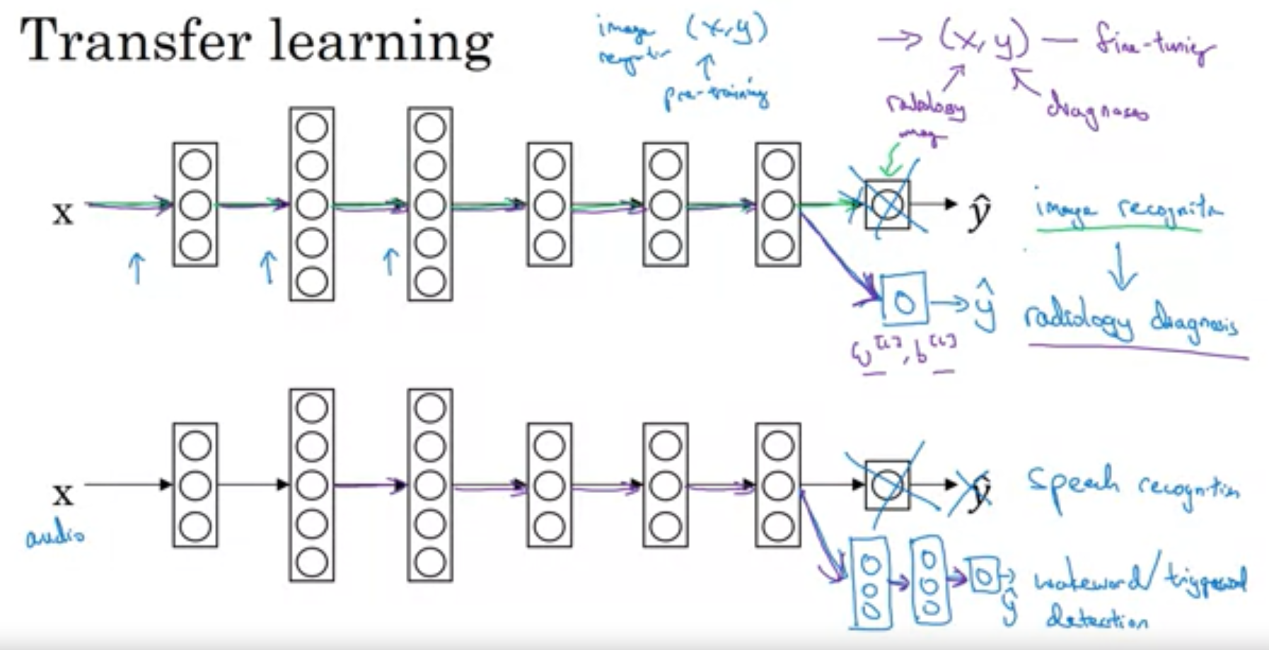
在准确率相互比较奇怪的情况下如上table可给出哪里出了问题（上图中原model方法完美）

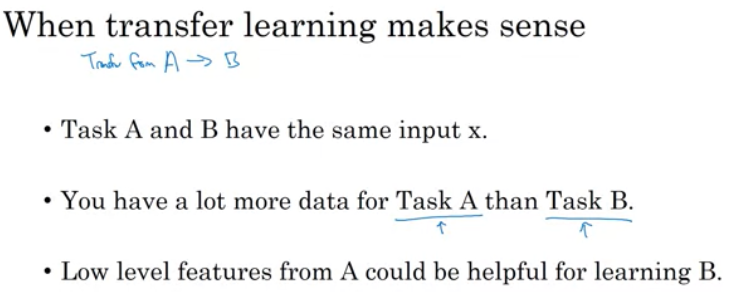
10. 尝试解决data-mismatch 问题：用artificial data synthesis

主要synthesize数据集之间最不同的特点（从error里找，比如噪音是dev/test错误率高的主要原因，那么可以把噪音加进其他data point里）

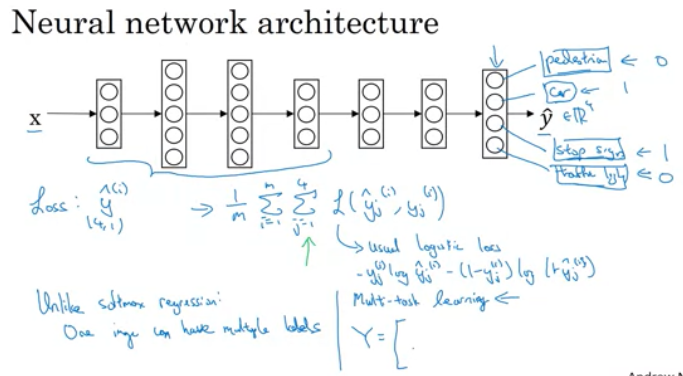
11. transfer learning

旧问题有很多数据，而新问题data较少，我们可以用旧neural net的一部分参数直接带进新问题



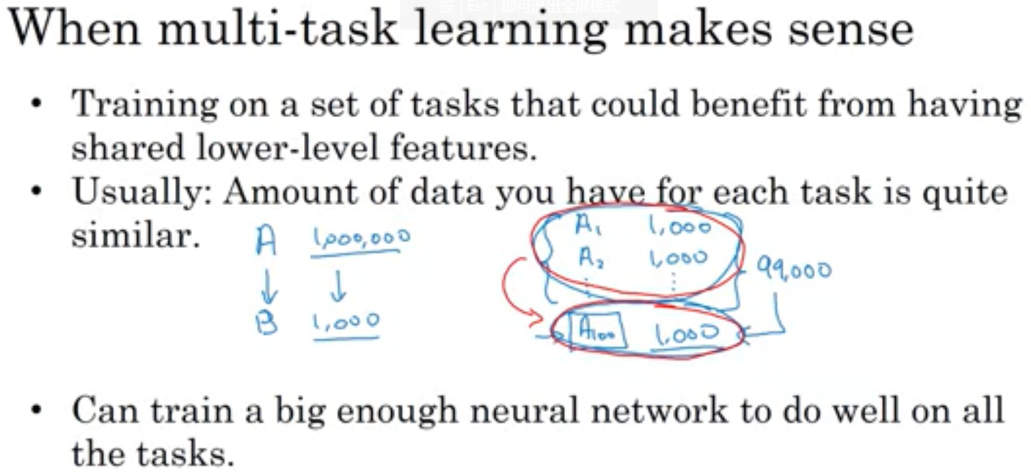
transfer A to B

12. multi-task learning：同时从多个任务中学习而不是只按顺序学习

一个输入data判断是否有好几个标签：loss用logistic的loss，只是把所有的都加在一起。

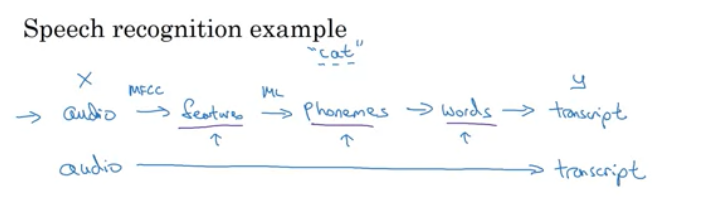
注意multi-task和softmax不同，softmax是互斥的，但上图中的multi-task不是互斥的，即；一个数据可以有多个标签，比如一张图里有行人汽车和红绿灯。

也可以对每一个标签都训练一个神经网络



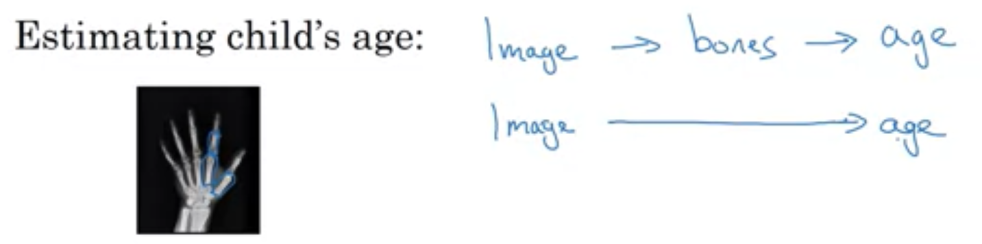
13. end to end dl:

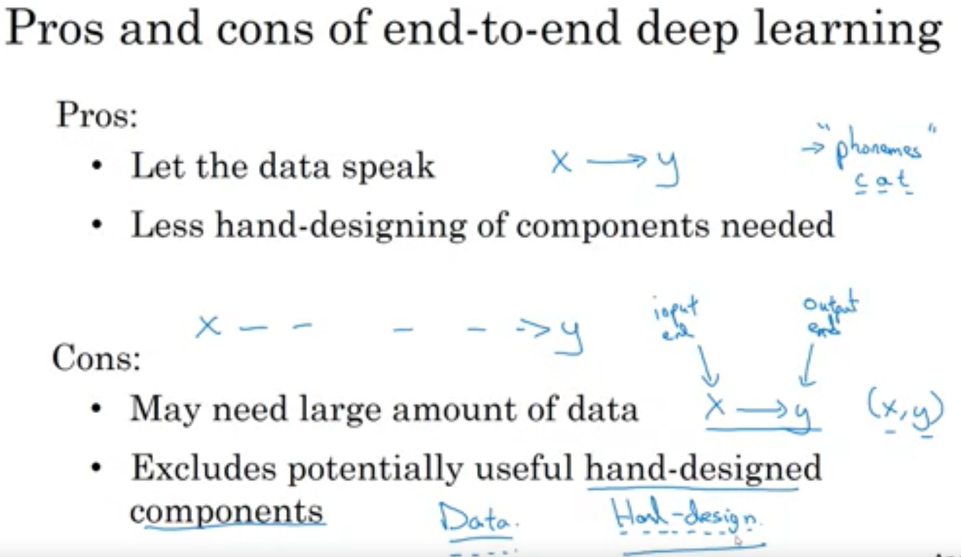
原来是把一个大问题分解为好几小步：即可能需要几个神经网络分别判断中间的步骤



而end to end指的是创建一个神经网络，把数据扔进去，直接得到最终想要的结果。这种做法的限制是可能没有那么多的数据。而分步解决问题在每一步中我们都可以寻求相关的数据。

比如人脸识别分为首先判断人脸位置，其次判断是否是这个人。人脸位置可以有很多数据，而是否是同一个人也可以有很多数据，两步的数据搜集是分离的。





如果有human-designed component injection 效果主要是可以注入一些人类自身的知识，免去了神经网络要从大量数据集里学习的过程