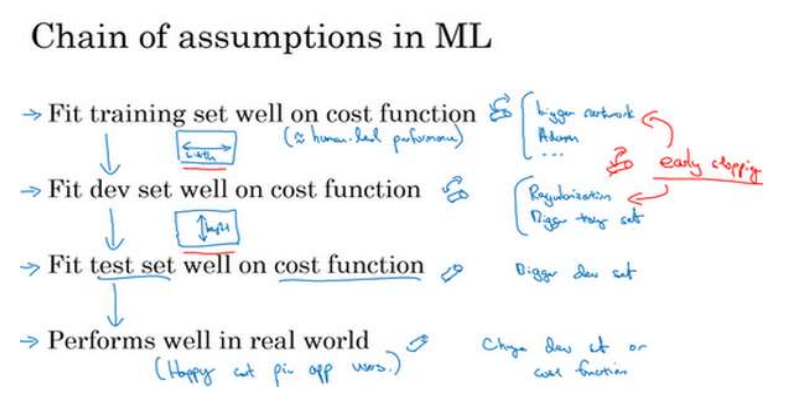
结构化机器学习项目

# 2 正交化

1. 调参依据参数的正交化原则. 彼此不相互干扰的参数.
2. 首先, 要保证在训练集上表现不错(通过某种指标)
3. 可以尝试调节参数,已达到在训练集收敛
4. 然后让网络在开发集(dev set)收敛(表现好). 需要调节(正则化)的参数.
5. 如果在测试集上表现不好, 需要回退到c)步骤, 使用更大的开发集来调试.
6. 如果在测试集上表现好, 在实际应用表现不好, 需要回退到d), c)步骤, 使用更大的数据集, 修改损失函数.



# 3 单一数字评估标准

Singal number evaluation metric

## 3.1 为了解决什么?

为了衡量(评估)一个手段(网络)的好坏, 为了更高效评估, 引入一个单数评估标准. 它能更高效评估.

## 3.2 有什么评估指标?

Precision 精确率, 查准率.

 其中, TP是正判正, FP是负判正.

Recall, 查全率.

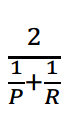
其中FN是正判负.

考虑如下情景:

1. A,B是两个”猫”分类器
2. A,B各自在P和R上有优势.
3. 从P和R上得不出哪个更好(更有优势).
4. 引入F1分数.
5. F1是P和R调和平均数
6. 描述P和R的组合效果.
7. F1分数是单一数字评估标准. 这样有助于(量化)那个分类器更好



F1分数.



有一个单实数评估指标真的可以提高你的效率，或者提高你的团队做出这些决策的效率

# 4 满足指标和优化指标

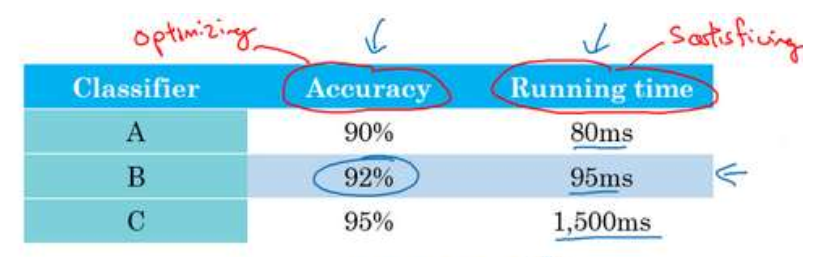
## 4.1 为了解决什么?

当难以用单一数字评估指标来衡量(评估)网络时. 当有多个指标需要考虑时.

引入满足指标和优化指标的思考方式.

具体讲,可以把一个指标作为优化指标, 其他N-1个作为满足指标.来衡量系统.

## 4.2 understanding 满足指标和优化指标



1. 对于A,B,C分类器. 有两个指标(大于一个,非单一指标)
2. Accuracy是优化指标, 含义是越大越好.
3. Running time是满足指标, 只要满足一定阈值即可. 只要满足time小于100ms, 至于80还是95ms,对用户体验不大.
4. **满足指标是先要达到的, 在满足指标下, 优化指标越好越好.**
5. 综上, 选择B分类器, 因为其处于满足Runningtime的满足指标下, 其优化指标最大.

# 5 训练/开发/测试集分布

## 5.1 为了解决什么?

为了让机器学习项目更高效收敛, 为了能够更好的接近预订性能指标. 需要小心考虑训练,开发,测试**集合的分布.**

1. 关注点不是说”训练/开发/测试”集这个主意.
2. 重点在于三者的样本分布.
3. 期望是有同一分布(样本呈现出相似分布).

## 5.2 分布

a) 让你的开发集和测试集来自同一分布

b) 将所有数据随机洗牌，放入开发集和测试集

# 6 训练/开发/测试集的大小

1. 传统的是70%/20%/10%. 但是在大数据时代不适用了.
2. 有 10,000 个例子就能给你足够的置信度来给出性能指标了
3. 现在流行的是把大量数据分到训练集，然后少量数据分到开发集和测试集.
4. 1 百万个训练样本，98%作为训练集， 1%开发集， 1%测试集(1万个样本)