**Machine Learning Step:**

**1. Define a model(function set)**

**2. Define goodness of function(Lose Function)**

**3. Pick the best function**

**Where does error come from?**

1. 較複雜的模型會有較低的Bias，但可能會over-fitting training data，導致variance較大

2. 較簡單的模型會有較高的Bias，無法fit training data

**那怎么知道现在是bias大还是variance大呢？**

1. 如果model没有办法fit training data的examples，代表bias比较大，这时是underfitting

2. 如果model可以fit training data，在training data上得到小的error，但是在testing data上，却得到一个大的error，代表variance比较大，这时是overfitting

**如何针对性地处理bias大 or variance大的情况呢？**

遇到bias大或variance大的时候，你其实是要用不同的方式来处理它们

1、如果bias比较大

bias大代表，你现在这个model里面可能根本没有包含你的target，可能根本就不在你的function set里

对于error主要来自于bias的情况，是由于该model(function set)本来就不好，collect更多的data是没有用的，必须要从model本身出发

**重新设计你的model**

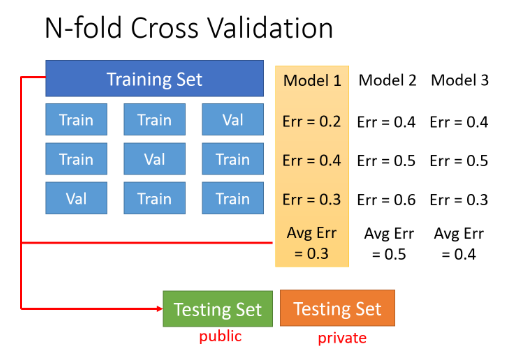
* 增加更多的features作为model的input输入变量
* 让model变得更复杂，增加高次项，比如原本只是linear model，现在考虑增加二次项、三次项…

2、如果variance比较大

* 增加data : 如果没有办法collect更多的data，其实有一招，根据你对这个问题的理解，自己去generate更多“假的”data
* Regularization(正规化): 就是在loss function里面再加一个与model高次项系数相关的term，它会希望你的model里高次项的参数越小越好，也就是说希望你今天找出来的曲线越平滑越好

調參==>調整model的參數，選擇model

訓練model ==> pick the best function, which have different weights in each feature



N-flod Cross Validation