动机：如何评估机器学习算法和模型

在机器学习的发展过程中，我们是先有了以下这些具体的算法和任务，再从中不断提炼出概念和指标来帮助我们评估机器学习算法和模型。在实际操作过程中，我们只需要明确的定义好我们任务可以了。

学习算法的基本定义

算法A，通过经验E，使得对任务T的性能指标P获得提升

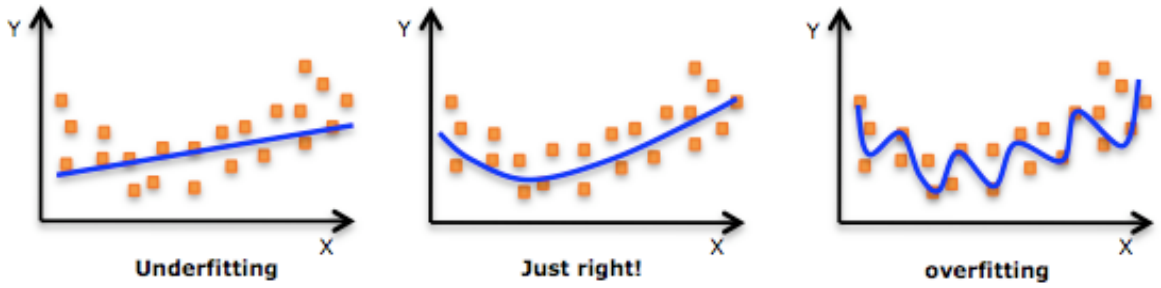
- T(task):
  - 分类:预测有限离散值
  - 回归:预测连续值
  - 序列/结构转换：机器翻译，图像$\longleftrightarrow$文字，word2vec
  - 数据的合成和降噪
- P(Performance):
  - Task-specific
  - $P_{test} \approx P_{generalization}$
  - $Error_{test} \leq Error_{generaolization}$
- E(Experience):
  - 无监督学习/监督学习
  - 联合分布分解成无监督学习 $p(\boldsymbol{x}) = \prod_{i=1}^{n} p(x_i \mid x_1, \ldots, x_{i-1})$

讨论

- Deep learning 优点在于在相对原始的数据上**自动的**、**更好地**提取出特征，所以在图像识别和语音处理等原先的特征提取水平较低的任务上取得更好的突破性成就。
- Deep learning 习惯于特征联系比较紧的对象，使得输入数据可以在它的**小邻域内做连续变化而不改变自身意义**

算法的学习能力

- 模型：算法：策略
- 独立同分布假设：训练误差的期望等于测试误差的期望,
- 容量：算法的拟合不同函数(分布)的能力
- 欠拟合： 不能学到训练集中的要求的分布
- 过拟合： 算法将训练集中与目标分布无关的特性也学习了

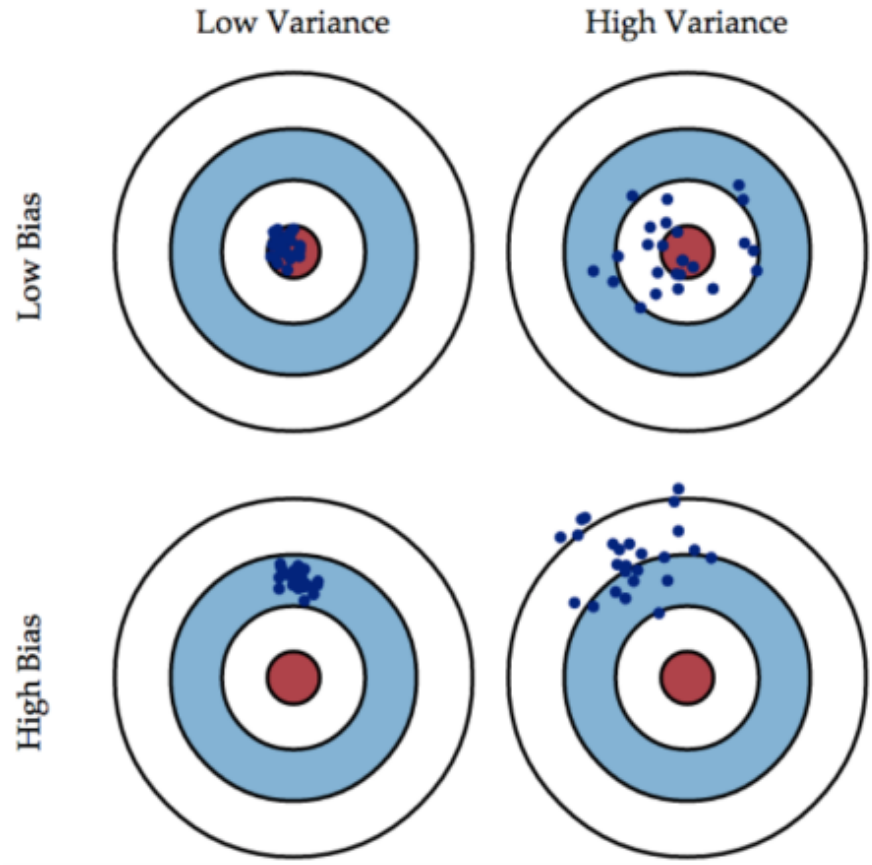
- 

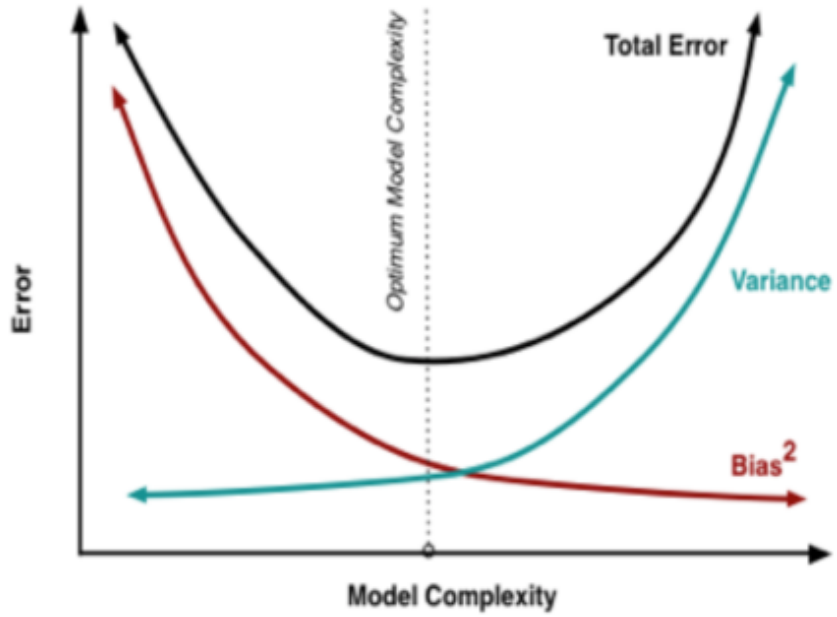
- $\sup\left|Error_{train}-Error_{generalization}\right|$: $\nearrow Capacity$, $\searrow Data$
- 没有免费午餐定理：
  - 所有可能的数据生成分布上都平均的意义下（平均考虑所有可能的任务），每一个分类算法在未知的数据点上都有相同的错误率。
- 也就是说不会有一个算法总是比其他算法要好。
  - 在实际任务中，对于特定的任务（特定的数据分布）就会有更适合的算法来学习。
- 机器学习的研究关注**什么样的算法**能在**我们关心的数据分布上**有更好的效果。
- **通过修改学习算法的假设空间来改变模型的容量**
- 正则化，通过给定正则化项（加入目标损失函数）来提供偏好（对那些情况对厌恶程度）

$$J(\boldsymbol{\omega})=loss_{train}+\lambda_{regularizer}$$

- 验证集， 重用训练集，保证算法学习过全部的训练集，也能通过不同的训练和验证过程，来测试算法（在特定的超参数下）的泛化性能。

准与确


Low Variance
High Variance
Low Bias
High Bias


Total Error
Optimum Model Complexity
Variance
Bias^2
Error
Model Complexity