

内容参考（<https://medium.com/towards-data-science/the-inductive-bias-of-ml-models-and-why-you-should-care-about-it-979fe02a1a56>）

归纳推理（Inductive Reasoning）



图 1

“一只戴着铃铛的斑点牛在草原上”，这可能是看到图 1 时的反应，然后，你观察到的是“戴着铃铛”、“斑点”、“牛”、“草原”，基于这些观察到的信息就会产生诸如“所有身上有斑点的牛都带有铃铛”，“草原上现在有许多头牛”，“所有牛都戴着铃铛”，“草原上只有牛”这些更加泛化的猜想（假设）。

那么，可以注意到，基于这些观察到的信息其实是可以做出很多上面那样的假设，这就是归纳推理一个重要的特性，即“有效的观察可能会导致不同的假设，并且其中一些假设可能是错误的。”

所以我们该如何选择一个且只此一个假设呢？根据奥姆剃刀原理“如无必要，勿增实体”，选择最简单的那一个假设“牛在草原上”，没有“戴着铃铛”、“斑点”这样的约束条件。奥姆剃刀原理也可以被视为最简单的归纳偏置，即如果有两个或多个原理，它们都能解释观测到的事实，那么应该使用简单或可证伪的那个，直到发现更多的证据。

机器学习中的归纳偏置

在许多机器学习任务中，我们总是试图用有限训练样本去估计无限的真实数据，例如在分类任务中，我们最常做的事就是从有限的子样本集中寻找一个可以

将所有样本进行完美分类的模型。那么在这个任务中，所观测到的事实就是数据集样本，基于此产生的一系列假设就是由从这个样本中学习到的所有可能的权重参数组成的不同的网络模型。每个模型对于新样本数据有会产生不一样的预测结果。

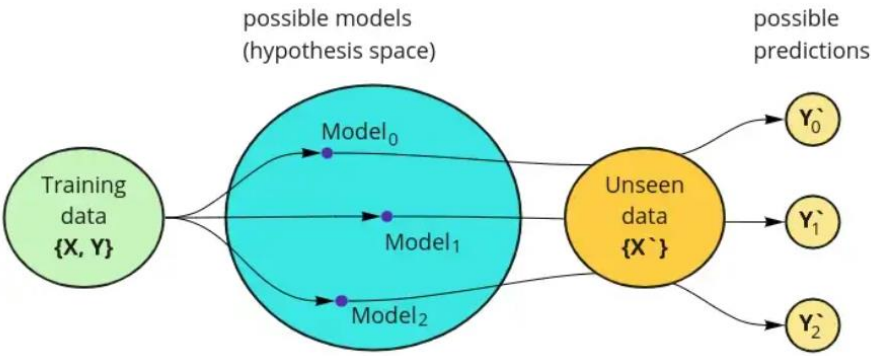


图 2

那么从有限的样本集中可以进行无限的假设。

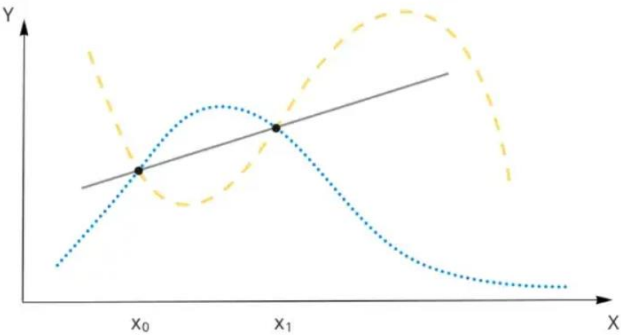


图 3

比如，图 3 是一个只含一个变量的函数，给出了两个已知点，那么黑色、蓝色、黄色三个曲线都可以拟合这两个点的数据，但是当出现一个新的数据点时（如图 4 所示），那么黑色的直线因为其与第三个数据点之间的偏差最小，所以看起来是最好的选择。

有了“偏差最小”这个先验条件，假设空间得以限制，也就是说假设空间中的一些先验条件，就是所谓的归纳偏置。

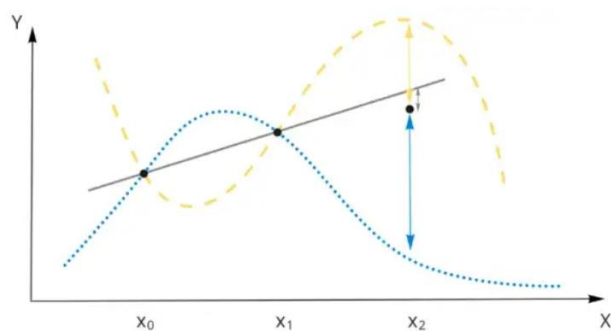


图 4

选择一个较强的归纳偏置，尤其是在我们数据量较少的情况下，可以帮助我们选择一个泛化能力较强的模型，我们的数据越少，我们就越需要一个强归纳偏置帮助我们的模型更加泛化。相反，当我们数据足够多时，归纳偏置可能会起到反作用。

那么，我们为什么通常会说“CNN 一般用于处理图像数据”，“RNN 一般用于处理序列数据”？这是因为这些模型有着不同的归纳偏置，例如 CNN 模型具有的归纳偏置都是局部性和平移不变性。局部性意味着相邻两像素之间相关性更强，平移不变性是通过权值共享的滑窗实现的，意味着图像不同部分会使用相同的处理办法，寻找的是特定的模式，也就是特征或者某种模式出现在图像中的任何位置时，其本质不变。

Vision Transformer 中的归纳偏置

Vision Transformer 中只有 MLP 模块里引入了局部性和平移不变性这两种归纳偏置，而在一开始将图像进行 patch 映射时，相当于将图像进行分块，无法携带二维位置信息，因此该归纳偏置就很弱，不过为了解决这个问题加上了位置编码，所有 patch 在空间上的相关性需要从一开始就进行学习。因此 Vision Transformer 中的归纳偏置和 CNNs 相比很少，所以这也就导致了当 ViT 在数据匮乏的情况下进行训练得到的模型泛化能力很差。