

数

是神经网络的本质

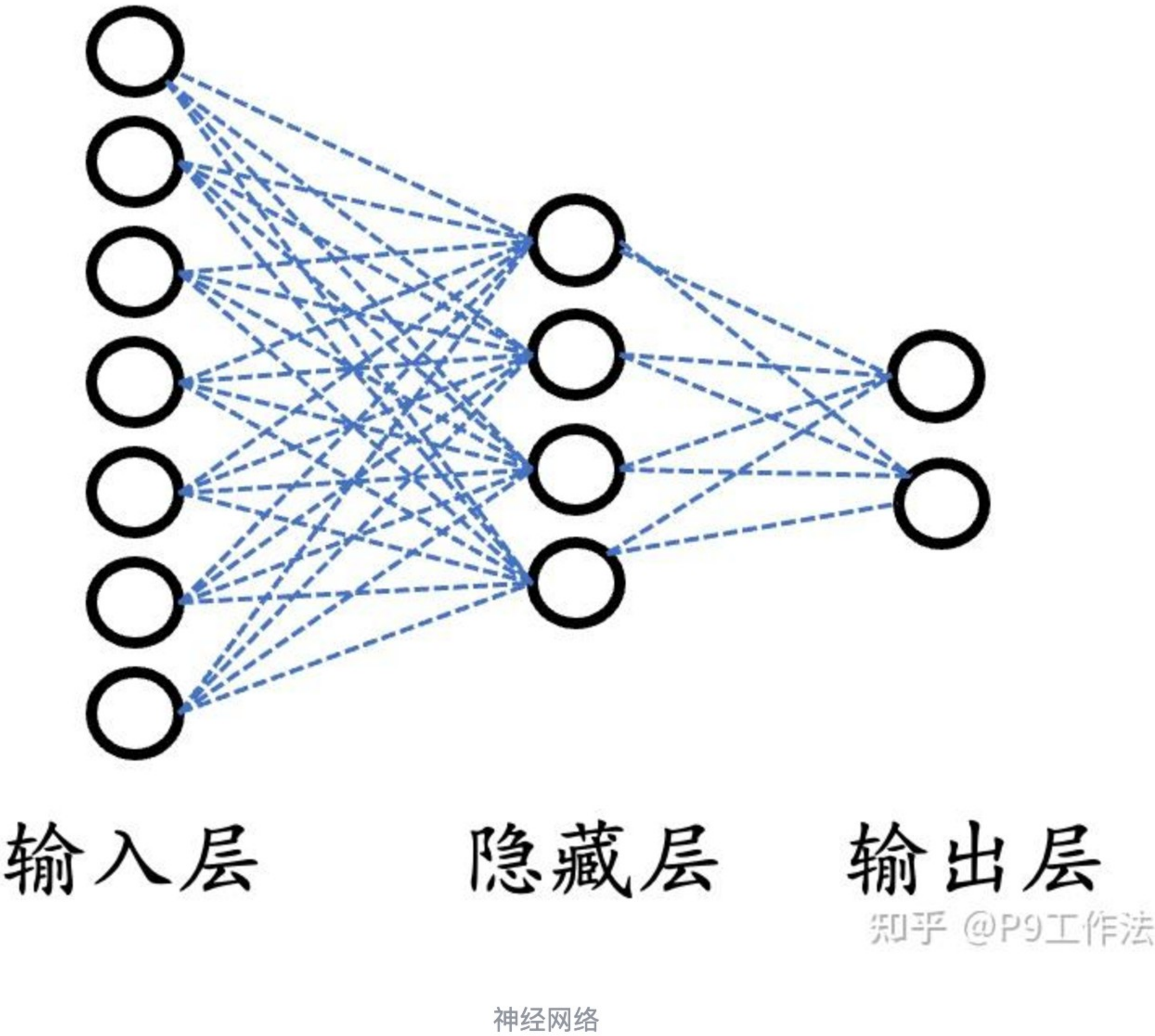
 **P9工作法**
分布式架构、上百人团队管理、全球支付实践与AI技术

已关注

23 人赞同了该文章

感性认知神经网络

自从大模型火起来以后，深度学习备受关注，而神经网络模型更是必须要深刻理解的。在了解其运行原理前，先来感性认知一下神经网络，如下图所示：



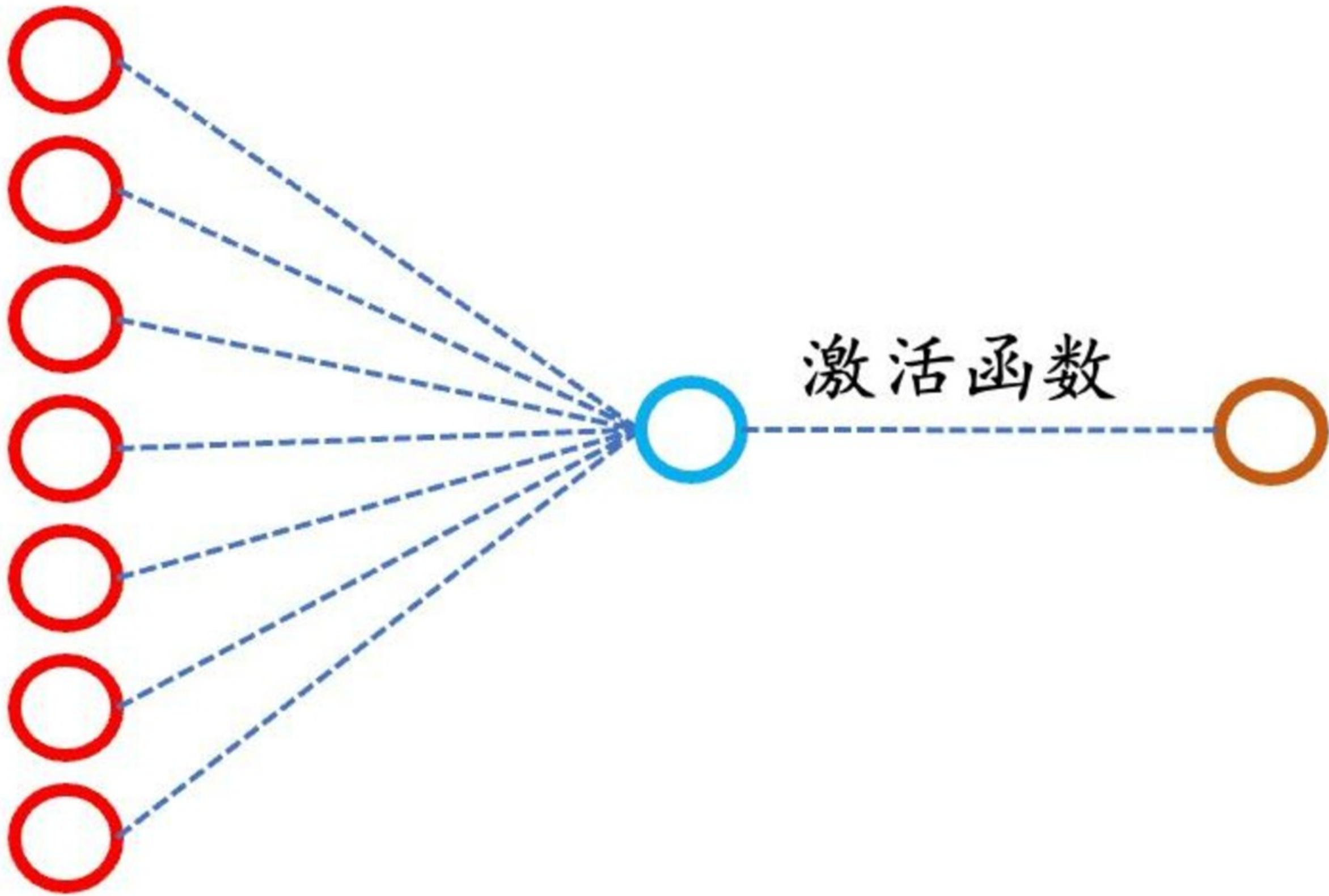
这是一个三层的神经网络，有输入层，隐藏层（当然输入层可以有多个），输出层。神经网络的使用过程可以想象为，输入一些信息，通过神经网络模型，就能够得到一个有效的输出。比如输入一张图片的信息到神经网络中，就能够输出这个图片是猫还是狗的概率。

稍微正式地总结神经网络：神经网络是一种机器学习方法，是受生物神经系统（大脑）启发的计算模型，能够从数据中学习复杂的模式，并用于分类、回归、聚类等多种任务。

神经元

神经网络由大量相互连接的人工神经元（或称为节点）组成，这些神经元被组织成层，包括输入层、隐藏层和输出层。

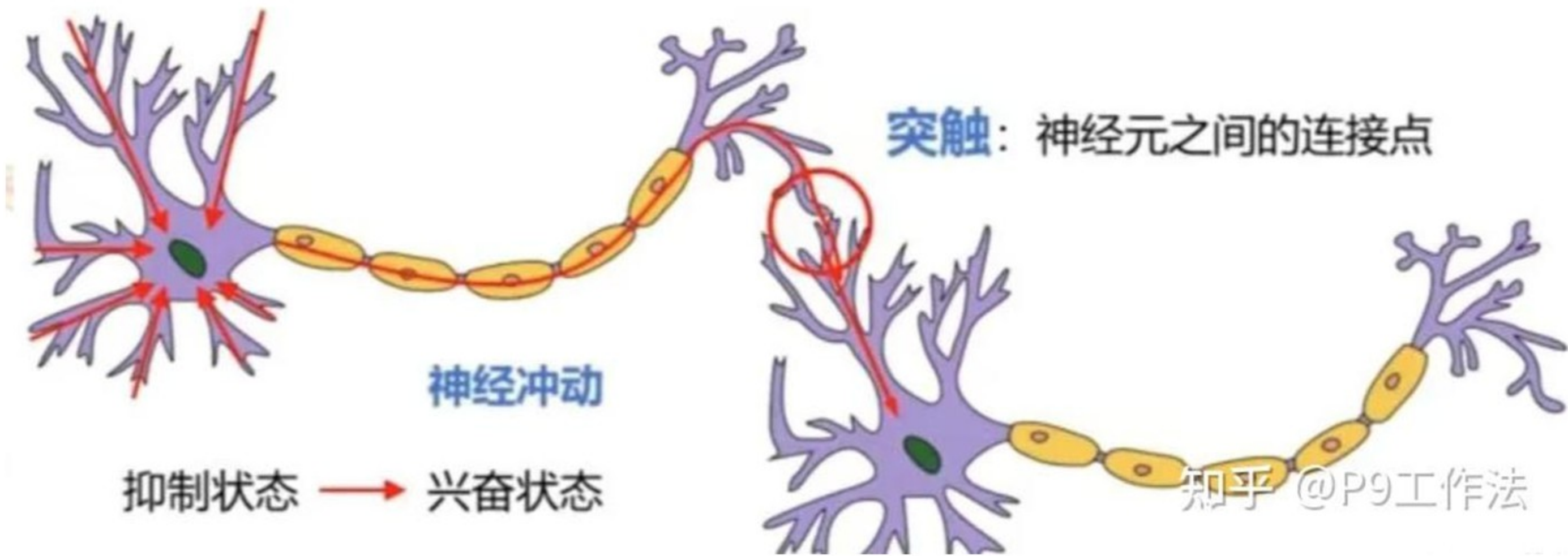
下图就是一个神经元，通过环境的各种信息输入（红色信号），到达蓝色的神经元，神经元得到刺激后会根据某种函数计算去看刺激是否到达阈值，如果到达阈值则往下传递到下一个神经元(a)。



知乎 @P9工作法

人工神经元

这个和人脑的工作原理类似，在外部刺激下激活某个神经元，然后向下再传递神经信号。



生物神经元

激活函数⁺

从大脑的神经网络的图片中可以看到，神经信号要被继续往下传递，是有一个激活/抑制的过程。模拟这个生物过程（其实后续还可以看到有很多地方都是受到生物特性的启发，比如drop out正则法等），使得对某些输入数据感知更敏感让某些数据感知变迟钝，于是就用了激活函数。比如常用的sigmoid函数⁺为： $f(x) = \frac{1}{1+e^{-x}}$

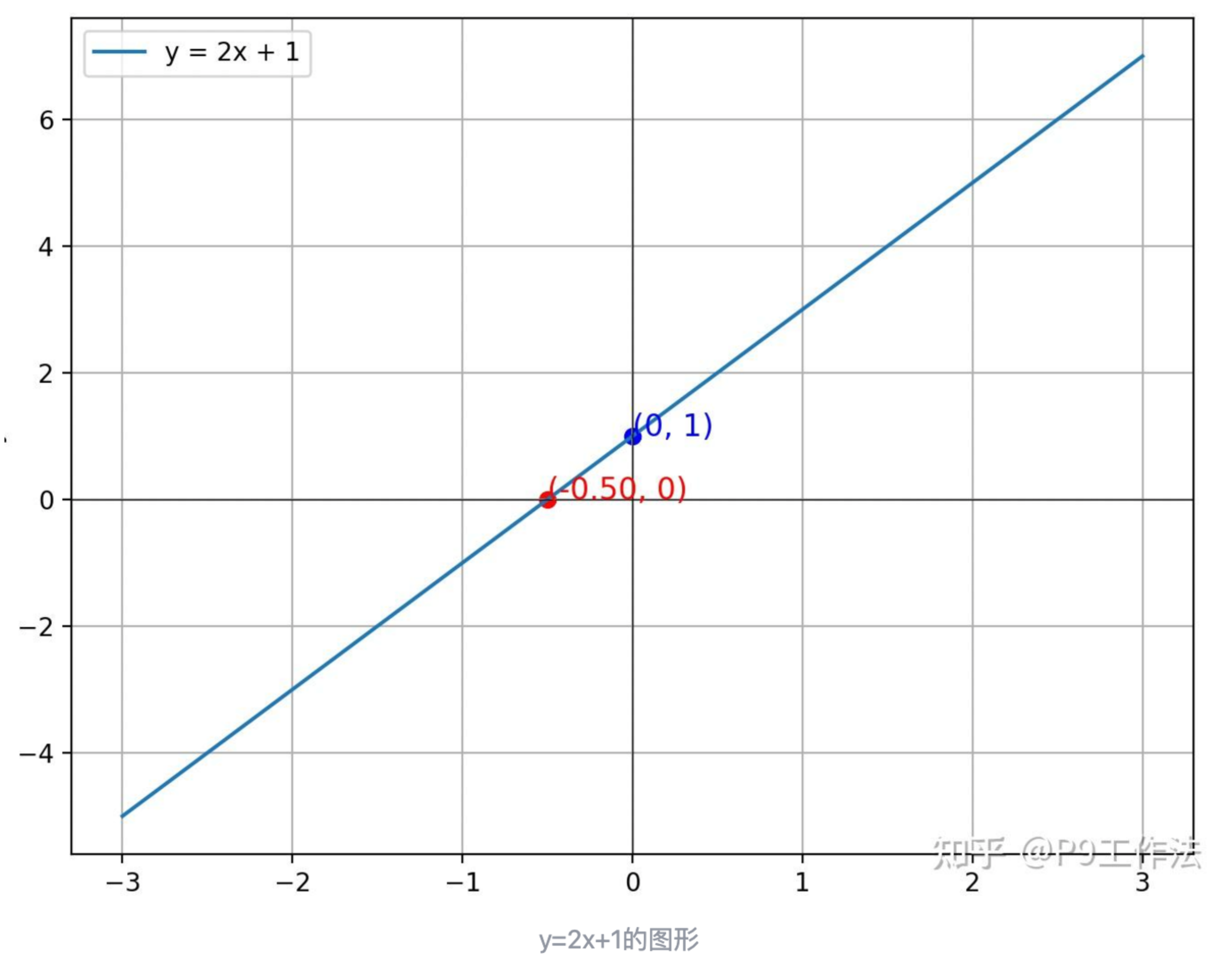
输入的信号通过这个函数运算，再往下一层传递。而神经网络中的激活函数也是有很多的，比如ReLU函数⁺，tanh函数等。

神经网络的本质

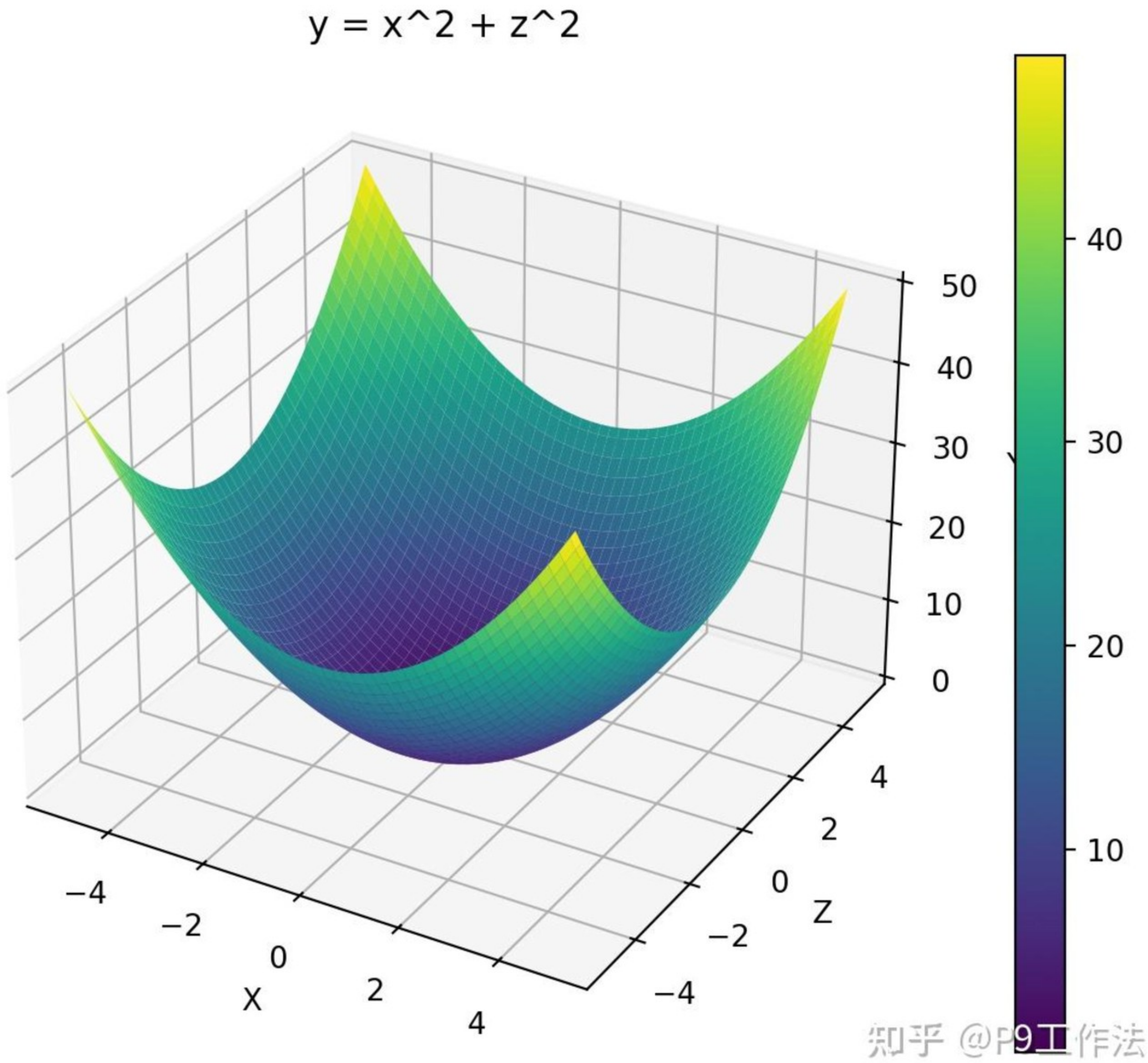
神经网络这个名字不明觉厉，会不自觉地想象难道是计算机能够模拟人的大脑运作么？且不说目前还没有完全搞清楚人脑的运行机制，即使搞清楚了该如何编程来实现？要搞清楚一件事还是得看透其本质，对于应用架构建模来说就是去以物理客观世界为锚定做抽象设计，而对于AI来说就必须回到数学，从数学书找到根基，才能够算是找到了本质。

从函数开始

首先来认识一下什么是函数：在数学书上的定义，函数是将一个对象转换为另一个的规则。例如 $y=2x+1$,



这就是一个函数，将每一个x变量映射为2倍+1的另外一个变量y。x取值为1，y取值就是3。这是一种简单的函数，更加复杂的函数可以是多个自变量⁺的，当然运算规律也有可能是平方而不是线性关系。



更多变量非线性的函数

按照函数定义，把映射的对象从数字变成其他的就很有意思，例如将羊肉变成羊肉串，这也可以用函数表达出来。

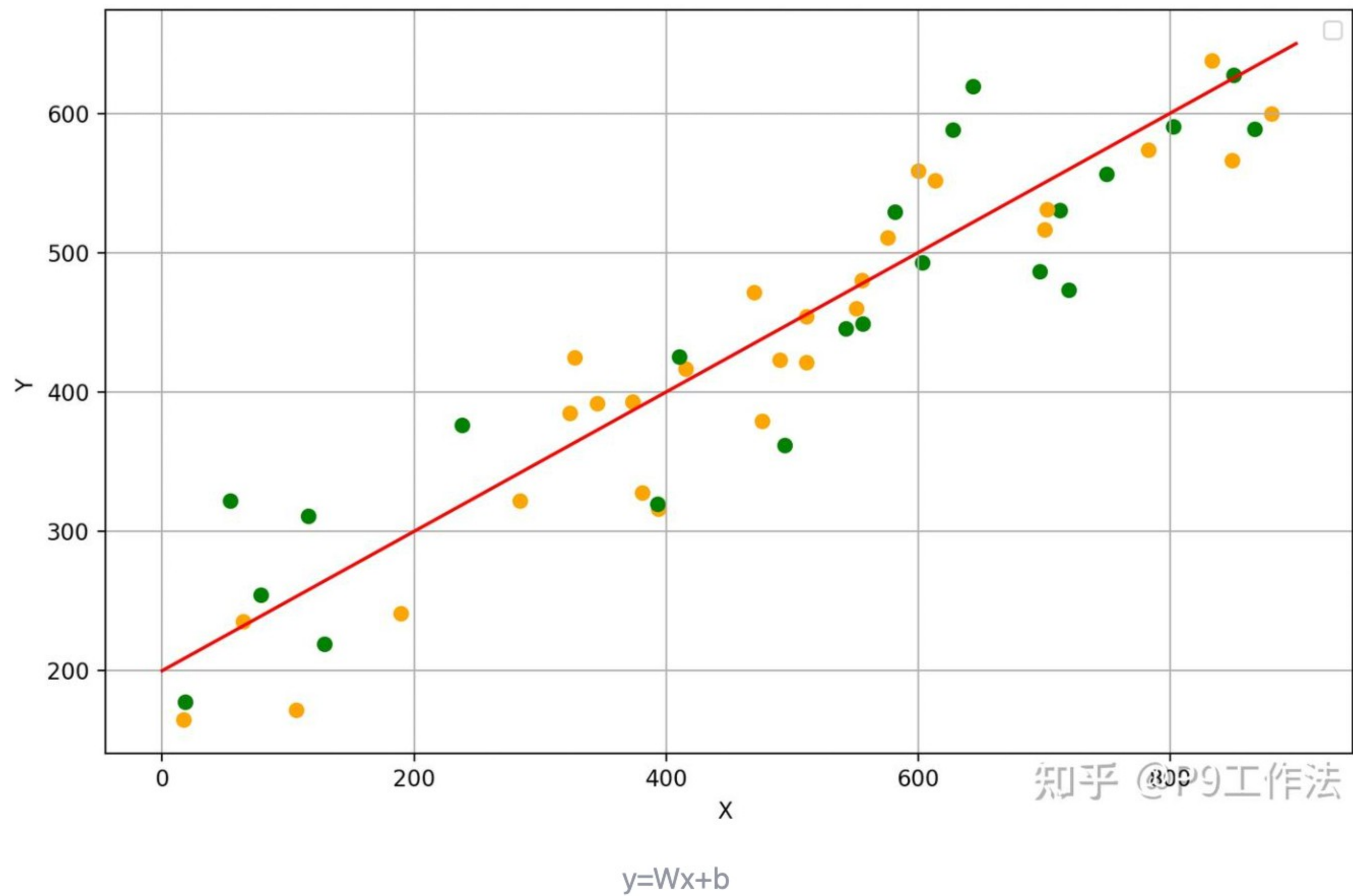
$$f(\text{羊肉}) = f_{\text{备料}}(\text{羊肉}, \text{腌制}, \dots) + f_{\text{调料}}(\text{羊肉}, \text{孜然}, \text{盐}, \text{味精}, \dots) + f_{\text{火候}}(\text{羊肉}, \text{时间}, \text{温度}, \dots)$$

，但是这个函数特别特别复杂，变量非常多，变量之间的关系也特别复杂。这也是人脑更高明的原因，一碗饭就能够给大脑充满能量烤出来还不错的羊肉串，但是AI要训练得到这个模型可就太耗费电力了。

直线拟合数据

上诉烤羊肉串的情况太复杂了，我们用预测房价的简化版本来举例。显而易见的，我们会把影响房价的变量列出来，比如：面积，年限，楼层，环境，小区物业，学区等各种各样的数据。

首先想到的就是简化变量，假设先考虑面积这个因素，函数写成 $y=Wx+b$ 。那么势必会得到这样一个图。

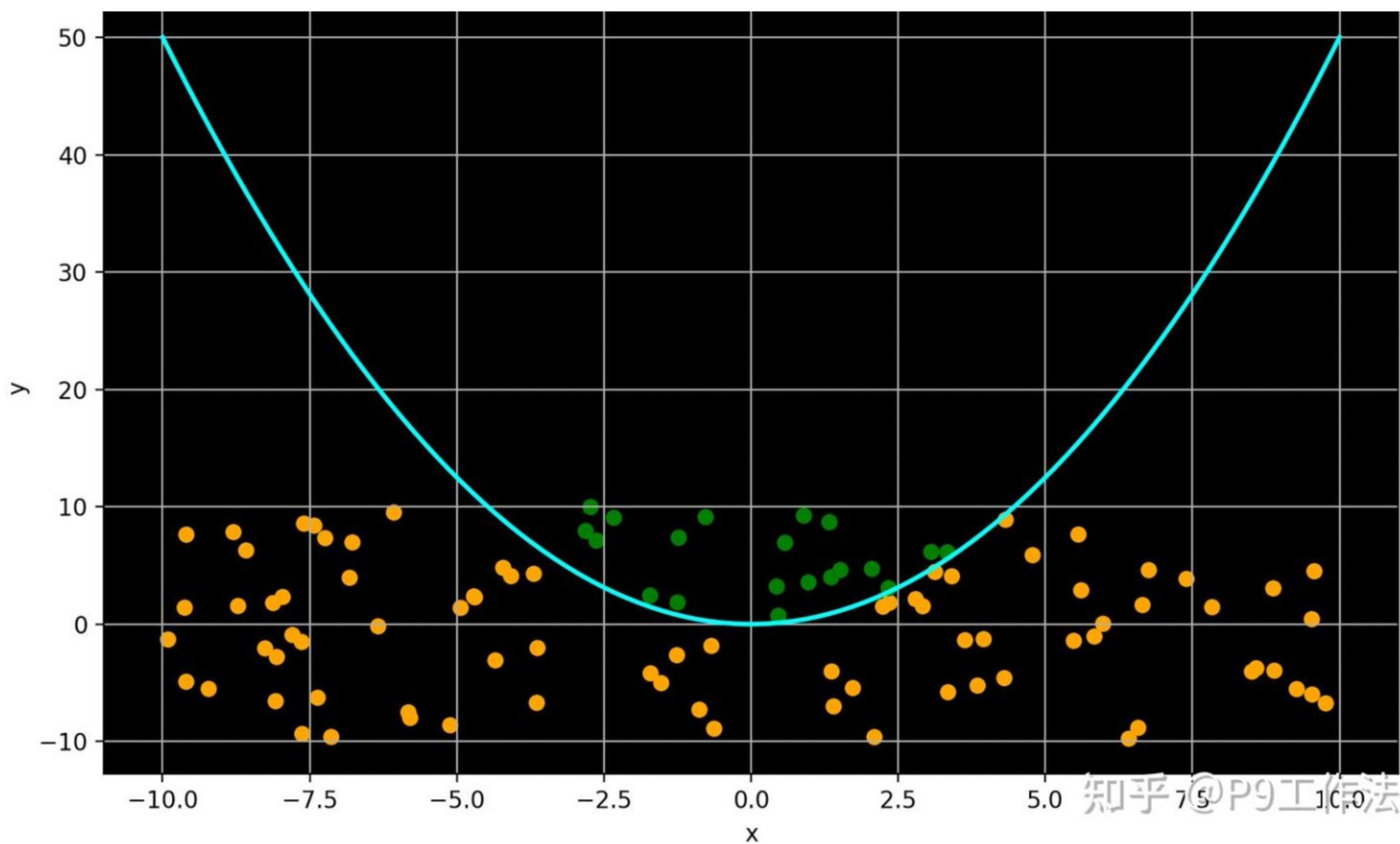


在图中，一条直线能够把数据区分为两堆，如果再来输入一套房的面积，大概是能够知道房价是落在哪一堆，也就是实现了放假的预测。这也叫数据的拟合，一个好的模型能够很好拟合数据。

非直线拟合数据

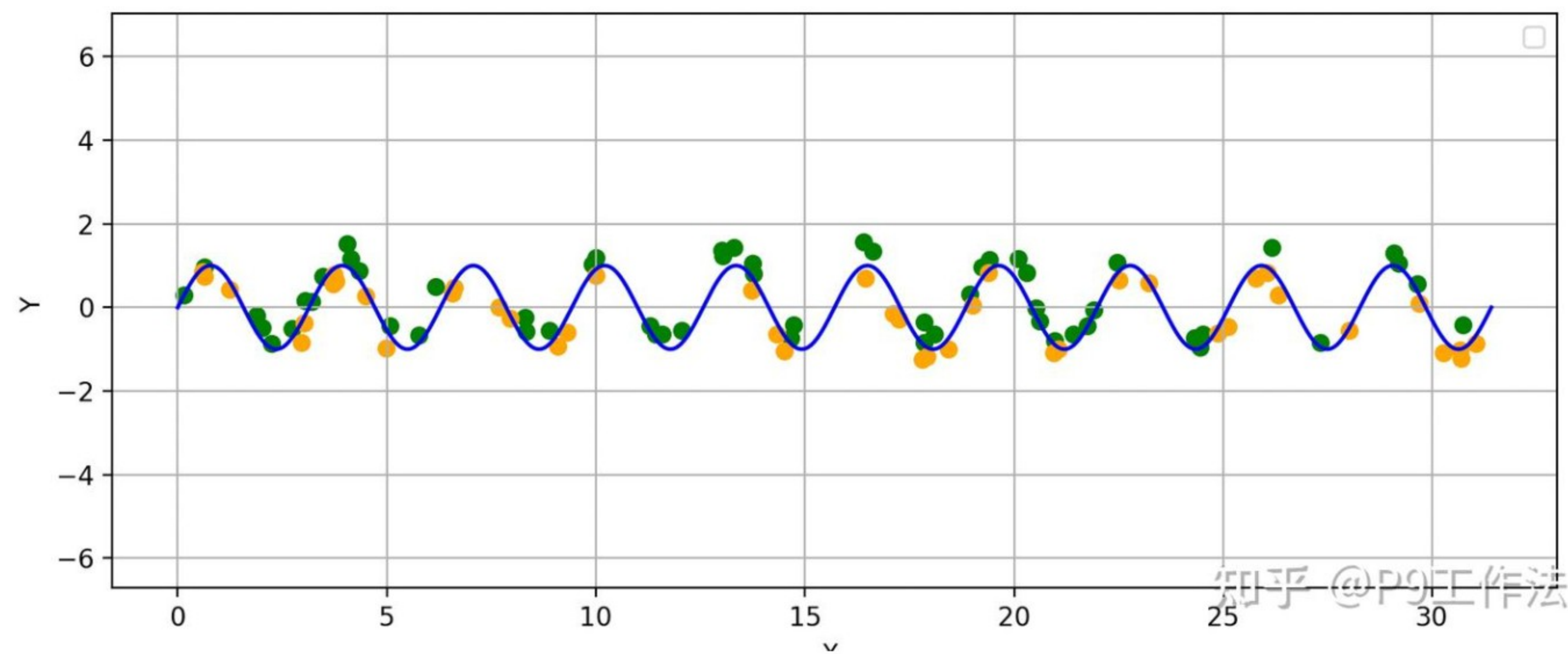
当然我们知道，上面的直线函数太过于简单了，现实生活中不太可能是简单的直线。也许数据复杂得需要曲线（而且直观感觉应该曲线会更靠谱），如下图所示的曲线：





曲线

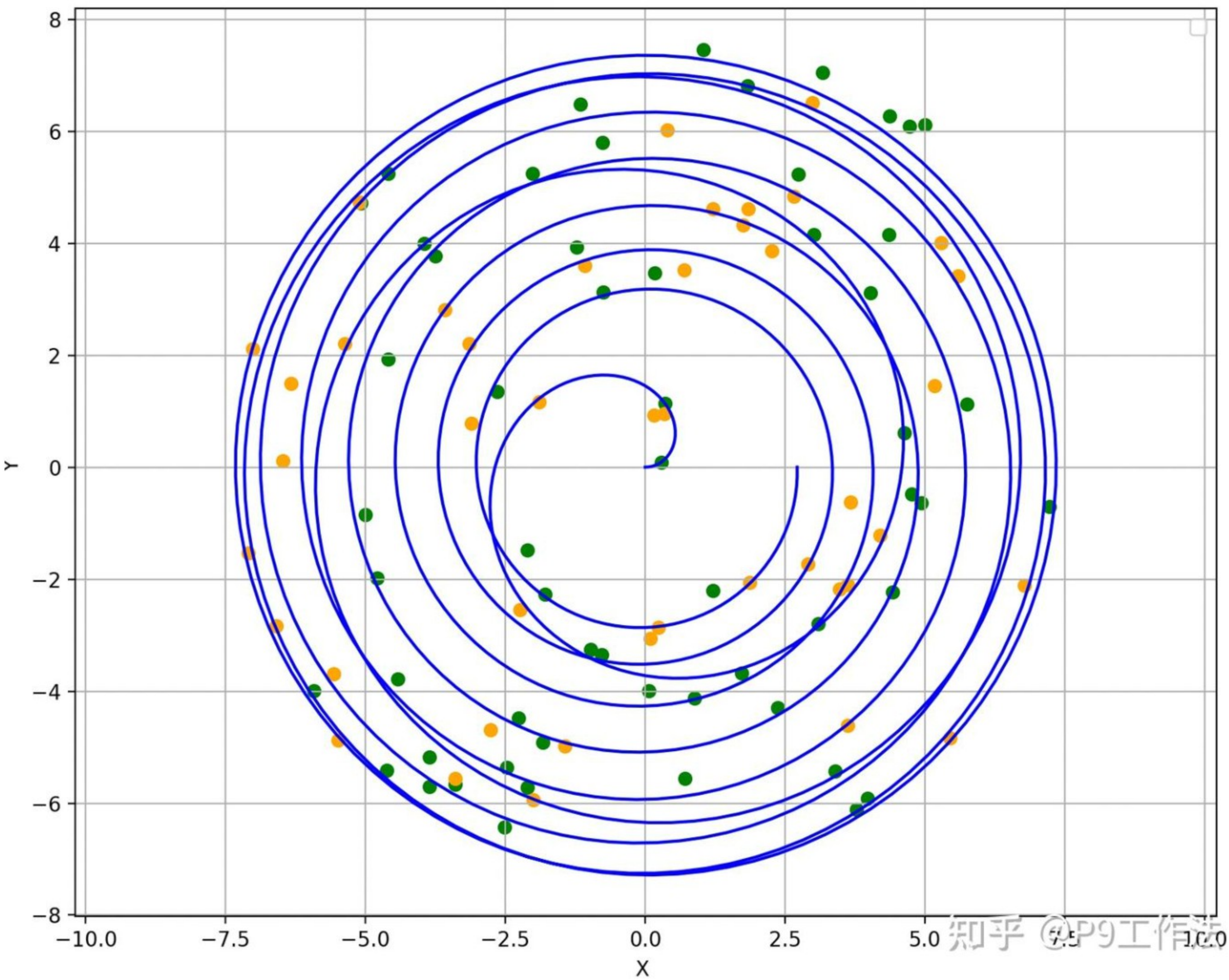
那如果数据分布越来越复杂，可能要找到符合数据分布的曲线也就会越复杂，如下图：



更复杂的曲线

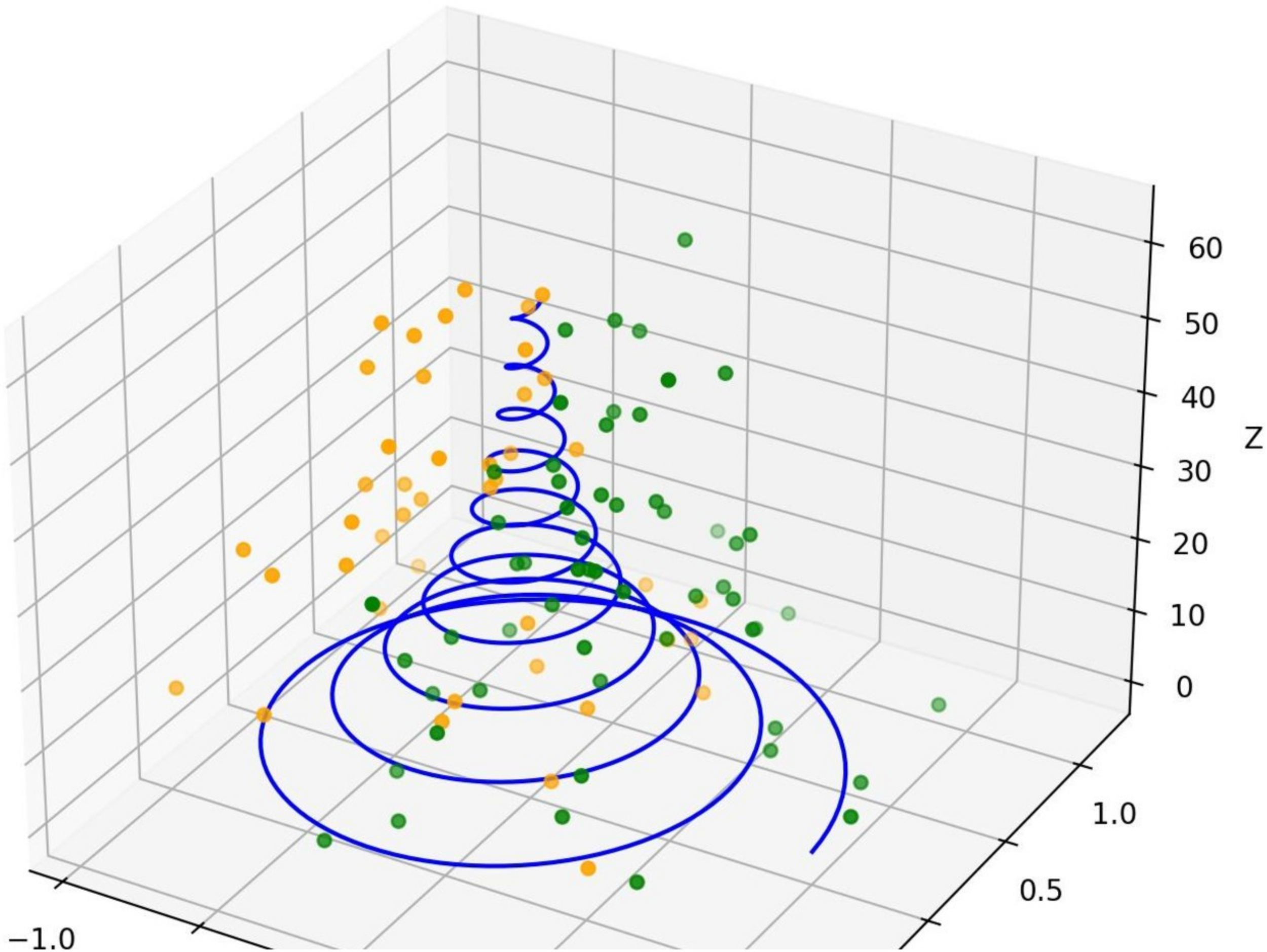
空间拟合数据

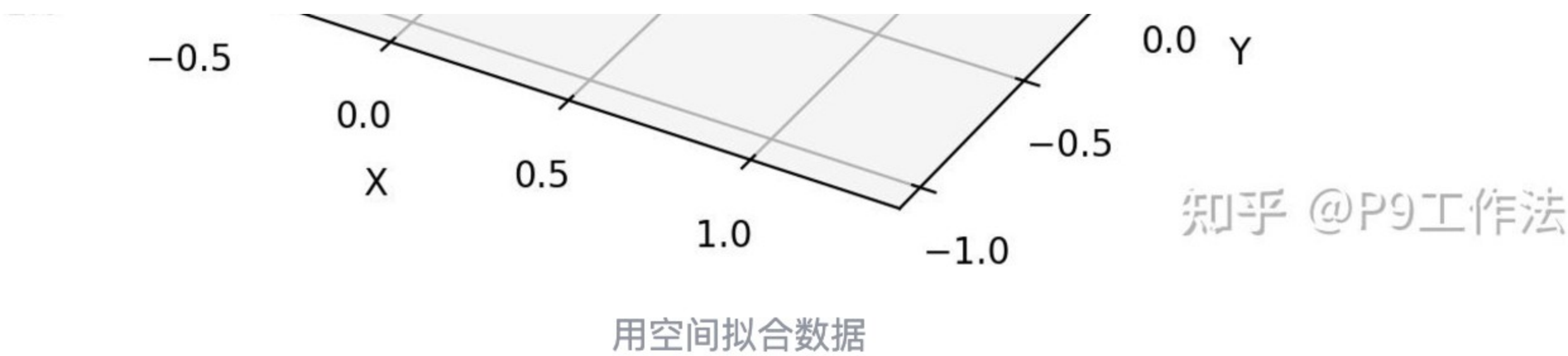
有时候甚至二维的都不够用，如下图的数据分布，曲线的拟合效果肯定很差。



曲线也未必是最好

拉升为立体图形+肯定效果会好很多，如下图所示：





从二维平面到空间会让很多问题好解决很多，而且**多维空间**在数学上也是非常容易表达和计算。当然空间也有很多种，有常规的欧式空间，也有**非欧空间**（希尔伯特空间等），不同的空间就可能更快解决某些问题。

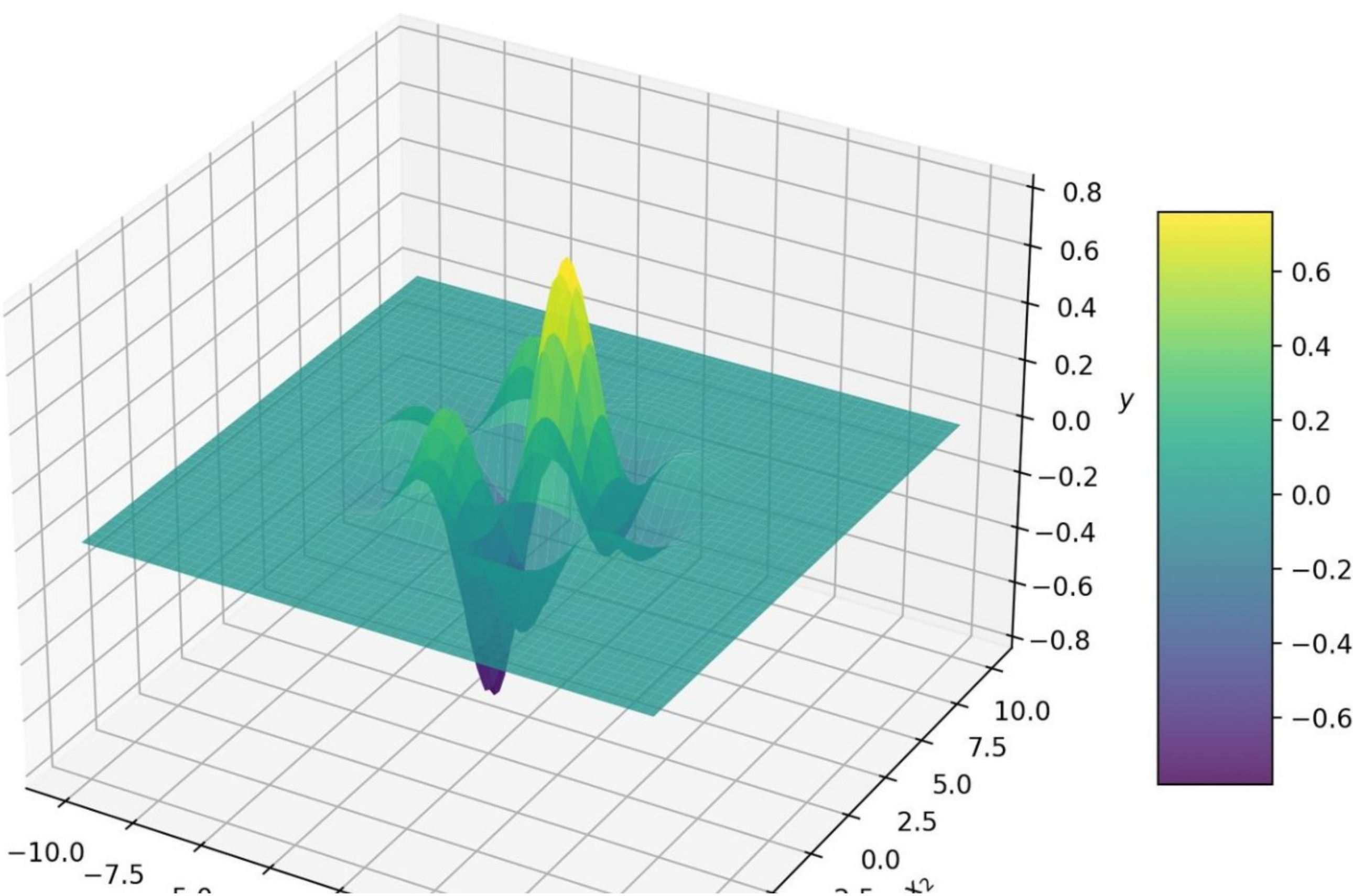
通用近似定理

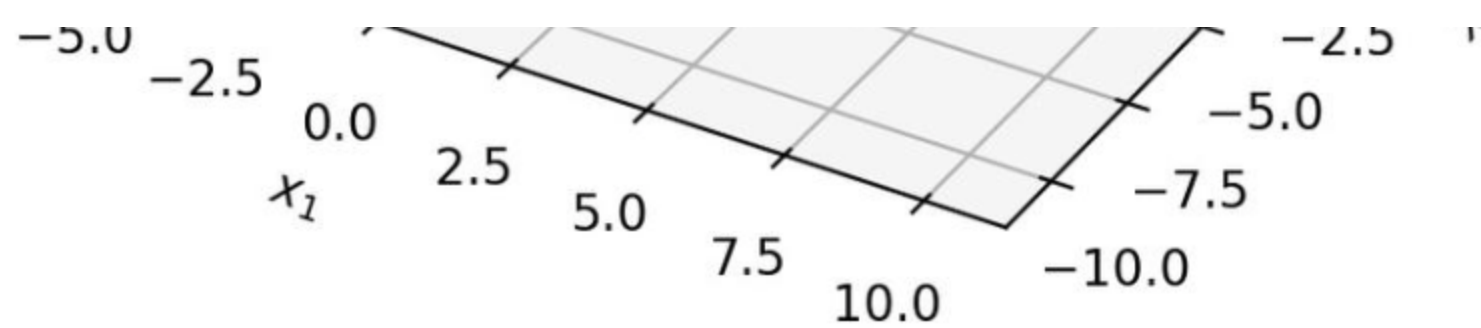
所以从数学角度来说，要想更好拟合数据，应该是需要更加复杂的函数才对。这一点也从实际经验来看也是这样，首先变量不太可能是线性关系，其次变量不可能只是房价一个，而是有很多个变量在起关键作用。

如果单是把上面列举的面积，年限，楼层，环境，小区物业，学区等因素列出来，这个函数至少要表达成： $y = W_1x_1 + W_2x_2 + W_3x_3 + W_4x_4 + W_5x_5 + b$ 。如果考虑非线性的关系，比如 x_1 是平方或者立方关系，或者把 W_1 换成cos函数（就是变成**复合函数**），可能才是更好的函数表达。

其实这也解释了为什么神经网络需要激活函数，因为通过激活演示去转换，才能够更好去拟合数据。

当神经元足够多时，也就是上面的函数是有足够多的维度，就能够做出来非常复杂的拟合图形（就是维度足够多，空间足够复杂）。数学家已经证明，只要神经元足够多（大于100个），任何复杂的曲线都能够被模拟出来，这叫做通用近似定理。





知乎 @P9工作法

空间来拟合数据

小结

所以神经网络的本质就是函数，通过数据训练得到更好的一个函数去拟合测试数据，并用这个函数去计算未来的数据。从这里可以得到如下几个要点：

- 1、AI与传统的应用架构有本质区别的，传统应用架构是设定参数和计算规则，用现有数据做计算得到。而AI的过程是数据训练（计算）得到参数，用参数再结合现有数据计算得到结果。
- 2、参数在这两个领域的略有差异，AI的参数是自动计算得到的，若真要把AI世界的参数与传统应用架构的建立起来联系，那应该是AI的超参数+与传统应用架构的参数对应起来。
- 3、从本质上上来看，AI就不是精确计算，因为函数系数是通过训练数据算出来的，未必适合于未来的数据。所以才会有数据过拟合+的现象（就是对历史数据拟合得很好，但是对真实数据计算就效果较差）。