从函数拟合开始

最简单的规律——简单线性回归

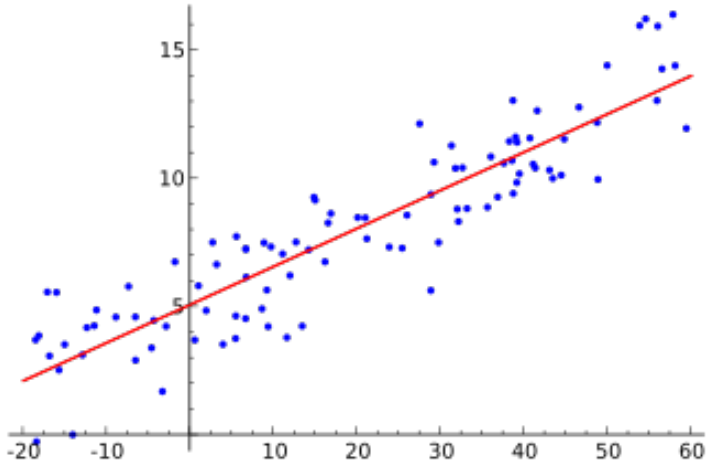

Figure 1: 线性回归示意图
图源：Wikipedia

虽然 Linear Regression 线性回归 的名字叫做" Regression 回归 "，但是事实上我更喜欢叫做 Linear Fitting 线性拟合 。它的目的是找到一条直线尽可能"贴近"数据点。在这一基础上，我们可以发现数据之间的规律，从而做出一些预测。不过这里有几个问题：

- 为什么要用直线？为什么不用曲线？
- 为什么要用直线拟合数据点？这有什么用？
- “贴近"数据点的标准是什么？为什么要选择这个标准？

我认为用直线的原因无非两点：一是直线 $y = kx + b$ 简单且意义明确，又能处理不少的问题。几何上直线作为基本对象，尺子就能画出；代数上只需要加减乘除，一次函数我们也很早就学过了。而它的思想一路贯穿到了微积分的导数并延申到了线性代数。二是许多曲线的回归可以转为线性回归（见后文）。

例如指数型的 $y = ke^{\alpha x}$ 取对数变为 $z = \alpha x + \ln k$，又如分式型的 $y = (\alpha x + \beta)^{-1}$ 取倒数转化为 $z = \alpha x + \beta$，从而归结为线性拟合。因此带着线性拟合经验再去考虑曲线会更轻松。

至于其意义：一是找到数据的规律，二是做出预测。拟合的系数可以用于测算数据之间的关系，斜率 k 表明输出对输入的敏感程度。一个经典例子是广告投放的 Marginal Benefit 边际效益 [1]，在一定范围内拟合收益与投入的关系，可以估算当前的边际效益，从而决定是否继续投放。而物理上，比值定义法定义的各种物理量，如电阻、电容等，最常用的测算方式都是线性拟合。例如测量电源输出的若干组电压和电流数据，并拟合出直线，斜率的绝对值是电源的内阻，同时截距顺带给出了电源的电动势，这样测得的数据就可以用于预测电源的输出情况。对我们所处的世界有定量的认识是科学的基础。可测量的数据和数学模型来描述、解释和预测自然现象是科学的基本方法，也是拟合的根本目的。

推荐阅读

[1] 边际效益：经济学概念，每增加单位投入，产出会增加多少单位

Least Squares

如果你想了解"回归"与" 最小二乘 “的含义：

用人话讲明白线性回归 *Linear Regression* - 化简可得的文章 - 知乎

https://zhuanlan.zhihu.com/p/72513104

如果你想阅读从求导法到线性代数方法的详尽公式推理：

非常详细的线性回归原理讲解 - 小白 *Horace* 的文章 - 知乎

https://zhuanlan.zhihu.com/p/488128941

如果你想详细了解了线性回归中的术语、求解过程与几何诠释：

机器学习| 算法笔记-线性回归（*Linear Regression*） - *iamwhatiwant* 的文章 - 知乎

https://zhuanlan.zhihu.com/p/139445419

多项式拟合

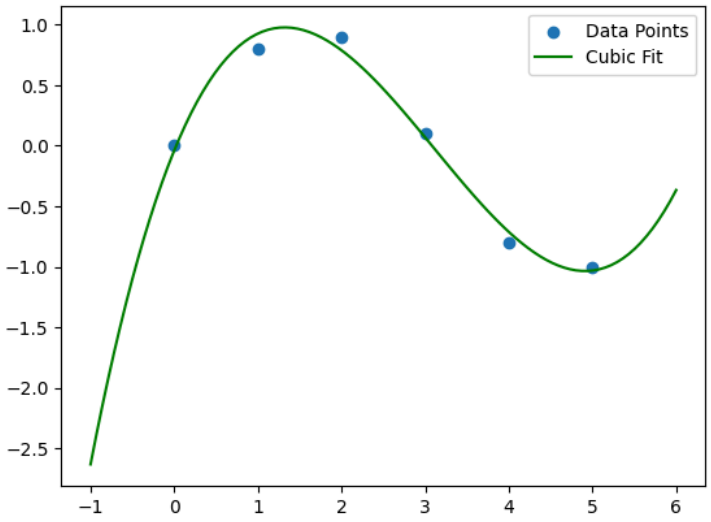

Figure 2: 多项式拟合示意图（图为 3 次拟合）
图源：GeeksforGeeks

线性拟合虽然很好，但是如果拿到了明显不线性的一堆数据，那么线性拟合就显得有些力不从心了。不过既然都是拟合，能做一次的那按理来讲也能做多次。 多项式拟合 Polynomial Fitting 就是这样一种思路，只是预测 $\hat{y}$ 从 $kx+b$ 变成了 $a_0+a_1x+...+a_mx^m$[2]，其中 m 是多项式的次数。而均方误差的表达式甚至几乎不用变，仍然是

$$\text{MSE}=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y})^2$$

只不过展开后是一系列的多项式项，待拟合的参数从两个变成了 $m+1$ 个。但是如果观察一下，这个式子仍然是一个（多变量的）二次函数，所以最小化的方法也是一样的。多项式自有多项式的好，能加的项多了，拟合的灵活性也就大了，误差显然会更小。然而与线性拟合相比，它虽然有 解析解 Analytical Solution ，但不再像线性拟合一样可以逐项明确说出意义，而是只剩下一堆矩阵运算把这些参数算出来。因此相比于记下公式，形成一个整体上的印象显得尤为重要。

上一小节中，我们从图像看到了这种拟合的几何解释，而多项式拟合也是相似的，还是从 $\boldsymbol{r}$ 的表达式入手

$$\boldsymbol{r}=\boldsymbol{y}-(a_0\boldsymbol{x}^0+a_1\boldsymbol{x}^1+...+a_m\boldsymbol{x}^m)$$

对比之前的表达式，当 $a_0,a_1,...,a_m$ 变化时，预测得到的结果 $\hat{\boldsymbol{y}}=a_0\boldsymbol{x}^0+a_1\boldsymbol{x}^1+...+a_m\boldsymbol{x}^m$ 也会在一个 $m+1$ 维的空间中变化，正如之前的平面，这个空间也是一个 $m+1$ 维的子空间。求最小模的 $\boldsymbol{r}$ 又回到了从点到子空间的垂线问题。虽然不得不承认：想象从一个高维的 n 维空间中向 $m+1$ 维的子空间做垂线确实有些困难，但是这多少离我们的几何直觉更近了一些。

系数的意义不那么明确了，但是误差下来了，这是好事吗？也不一定，灵活性的另一面是潜在的 过拟合 Overfitting 。前文中做线性拟合的时候有一个重要的假设是测量得到数据带有一定的误差。拟合的直线滤去了大部分的误差，留下了重要的趋势。但是如果灵活性太高，拟合的多项式会过于贴合数据，甚至把误差也拟合进去了。即使在给定的数据上做到了很小的误差，预测新数据的能力却可能会大打折扣。

[2]记号说明：虽然习惯上幂次从大到小排列，但是为了下标和幂次的统一性，所以这里选择从常数项到最高次项排列

拿做题打个比方：使用直线拟合明显不线性的数据是方法错了，只能说是没完全学会。但是用接近数据量的参数来拟合数据，留给它的空间都够把结果"背下来"了，捕捉到了数据的细节，却忽略了数据背后的规律，化成了一种只知道背答案的自我感动。在几道例题上能做到滴水不漏，但是一遇到新题就束手无策。

举个例子，在下面这个数据集上试图拟合，我们在二次函数 $y = 0.25x^2 - x + 1$ 上添加了标准正态分布的噪声，即实际上 $y = 0.25x^2 - x + 1 + \mathcal{N}(0,1)$[3]。

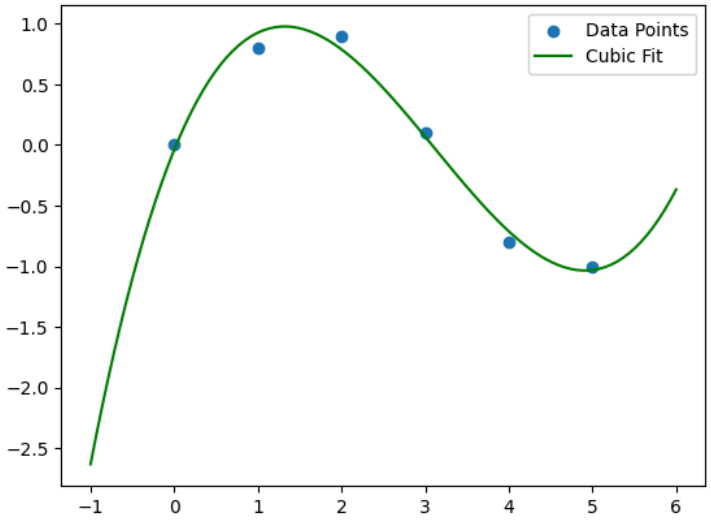

Figure 3: 多项式拟合示意图（图为 3 次拟合）
图源：GeeksforGeeks

那么现在我们来试试用不同次数的多项式拟合这个数据集。不难看出线性拟合的线与数据点还是相差不少，因为它没能提供可以制造数据"弯曲"形状的项，它没能捕捉到数据更加复杂的趋势，这种现象称为 欠拟合 (Underfitting)。2 次曲线的效果几乎和真实曲线一样，即使提升到 3 次也没有太明显的改变，它们拟合的效果都还算好。

[3] $\mathcal{N}(0,1)$：表示一个服从标准正态分布的变量，均值为 0，方差为 1