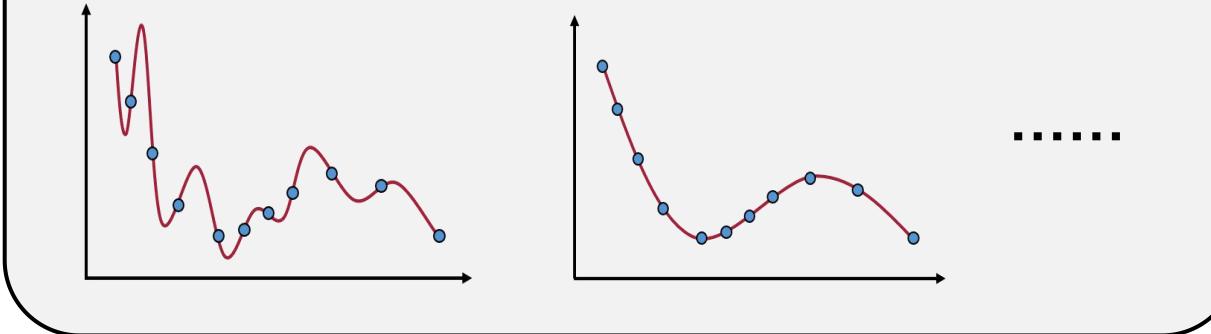


训练误差=0的集合：



参数量 > 样本数



多解

↑  
训练

泛化性好的解

实际上，损失函数无显式正则项：

$$L_{\theta} = \frac{1}{n} \sum_{i=1}^n (f_{\theta}(x_i) - f^*(x_i))^2$$

核心问题：神经网络的**隐式偏好**是什么？**机制**是什么？