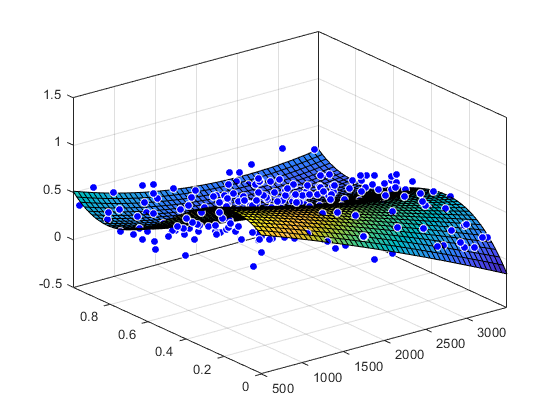
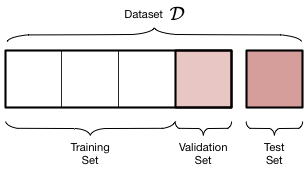
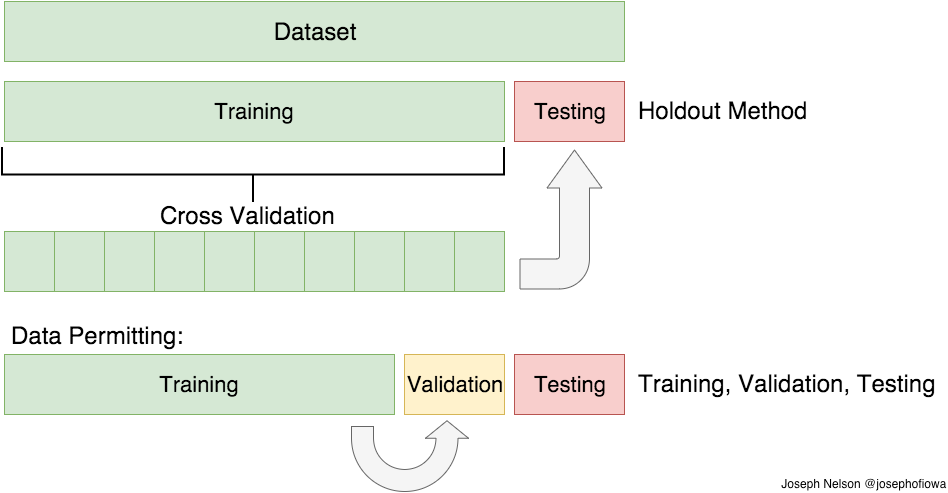
**欠拟合、过拟合**



当参数增加一个，模型变得复杂，数据集需要增加一个数量级。随着参数增加，导致所需数据急剧增加的过程，叫做维度灾难（Dimension curse）

发生过拟合、欠拟合的，一般是来源于样本数量和模型复杂程度之间的关系。

|  |  |
| --- | --- |
| **过拟合Overfitting** | **欠拟合Underfitting** |
| 训练数据占总体数据过少 | 训练数据占总体数据过多 |
| 模型过于复杂 | 模型过于简单 |
| 采样过于不均衡 | 采样过去全部 |
| 没有正则化 |  |
|  |  |



划分数据集：

* training：0.8，testing：0.2；
* training：0.7， validation：0.1， testing：0.2；
* cross validation（交叉验证）：可以充分使用有限的数据