**过度拟合**(Overfitting)是指对训练数据建模过好。当模型学习训练数据中的细节和噪声，对测试数据的性能产生负面影响时，就会发生过拟合。

**原因：**

1. 训练时间过长，模型学习到了训练集中的噪声信息
2. 模型过于复杂（例如神经网络的隐藏层太多）
3. 训练数据集噪声太多

**防止overfitting的方法：**

1. 用重新抽样的方法来测试模型精度（k-fold cross validation）.
2. 加验证集
3. 加Ln 正则化表达式
4. 剪枝（对于非参数模型，如决策树等）

**欠拟合**（underfitting）指所训练的模型既不能对训练数据进行有效拟合，也不能推广到测试数据集。

**原因：**

1. 模型过于简单
2. 训练数据集不够
3. 训练时间不够

**防止underfitting的方法：**

1. 增加模型复杂度（有一个最佳点，不能无限增加复杂度）
2. 增加训练数据集
3. 增加训练时长
4. 通过特征选择来精简特征变量数（特征变量越多需要越大的训练数据集）