1. Summarize the reasons of overfitting and underfitting：

解释过拟合和欠拟合之前，首先声明什么是训练误差什么是测试误差。在机器学习中，学习器的实际预测输出与样本真实输出之间的差异为误差。具体的误差可以分为训练误差和测试误差两类。训练误差是指，学习器在训练数据上的误差，也叫经验误差；测试误差是指学习器在新样本上的误差，也称泛化误差。训练误差描述的是输入属性与输出分类之间的相关性，能够判定给定的问题是不是容易学习的问题。而**测试误差则反映了学习器对于未知测试数据集的预测能力**，通常，我们都期望获得测试误差较小的学习器。

在学习器根据样本数据来拟合真实模型的过程中，如果一味追求训练误差的降低，学习器可能会把一些数据的特性当做所有数据的普遍性质，这样拟合出来的模型就会导致学习器泛化能力下降。最简单的例子，如果在"判定给出的动物是不是狗"的时候，样本中包含了哈士奇，而学习器把所有哈士奇的特征（皮毛颜色、耳尖等）作为衡量一个动物是不是狗的特征时候，可能在测试时候就会把一条金毛拒绝。这种就是典型的过拟合现象。通常过拟合是由于学习时，模型产生了过多的参数导致的。因而产生**过拟合现象**主要由以下一些原因导致：

1. **模型参数设置过多或训练模型过度（模型过度复杂）**
2. **训练数据有噪声**
3. **训练数据不足**

与过拟合对应的就是**欠拟合**，如果说过拟合是由于学习能力太强，那么欠拟合就是由于学习能力太弱，例如学习器把四条腿的猫的图片当成狗就是欠拟合的结果。导致**欠拟合的原因**主要如下：

1. **模型参数过少、模型简单**
2. **训练数据特征较多（不可省略）**

3. 训练不足

Writing down three sceneries that machine learning has been used now.

机器翻译

图像识别

语音识别

Come out with three new sceneries with which machine learning may be applied.

职业规划：根据性格特征、体质特征等，来给定一个人最适合的职业发展道路。

健康预测：给一个人的生活习惯，以及生理特征数据，来对一个人潜在的疾病进行预测。