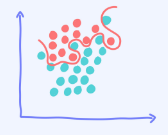
1. 解释监督学习，非监督学习，半监督学习的区别，并分别举例。

* 监督学习：数据集有输入和输出，通过输入数据和输出数据之间的关系构建一个映射关系，即函数的对应关系（例：x,y的对应关系y= wx +b，其中w,b为参数，x为输入数据特征，y为输出值或标签）。例如回归算法，就是通过建立输入输出的关系，训练模型后根据新的输入数据预测输出值，线性回归输出为连续值，逻辑回归为离散值。
* 非监督学习：数据集只有输入值，直接使用输入数据来进行建立模型，主要根据输入数据的相似度来进行对数据的归类。例如，K-means聚类算法，对数据集进行聚类。
* 半监督学习：将监督学习与非监督学习结合，数据集种少量数据有标记即有输入与输出的对应关系，但大部分数据是没有标记的只有输入值，综合两类数据来构建合适的分类函数。目前只接触过监督与非监督学习，对半监督的应用未知，查阅资料也没有给出能理解的例子。

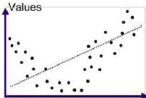
1. 解释模型过拟合，欠拟合，正则化，泛化能力，并分别举例。

* 过拟合：指构建的模型能很好的捕捉数据的特征并拟合数据，在训练数据表现得很好，而在测试数据表现得很



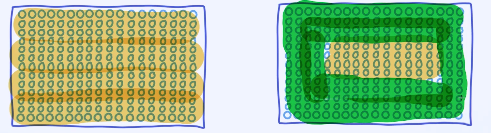
如图所示就是过拟合的现象。解决办法：增加数据量，使用简单的模型，加入正则项

* 欠拟合：是指构建的模型不能很好的捕捉数据的特征，在训练数据的时候就表现很差，不能很好的拟合数据



如图所示就是欠拟合的现象。解决办法：增加特征数量，减少正则化参数。

* 正则化：就是对误差函数加上约束，使参数w,b可行解变少，防止模型过拟合。也就是将先验知识加入到模型里，更好的生成预期的模型。



如图所示，左边是未加入正则，w,b的解空间，右边是加入正则后的w,b的解空间

* 泛化能力：是指训练的模型对新数据样本的适应能力。如果模型对新样本数据适应好说明模型的泛化能力强，反之很差，例如过拟合的模型泛化能力就很差。