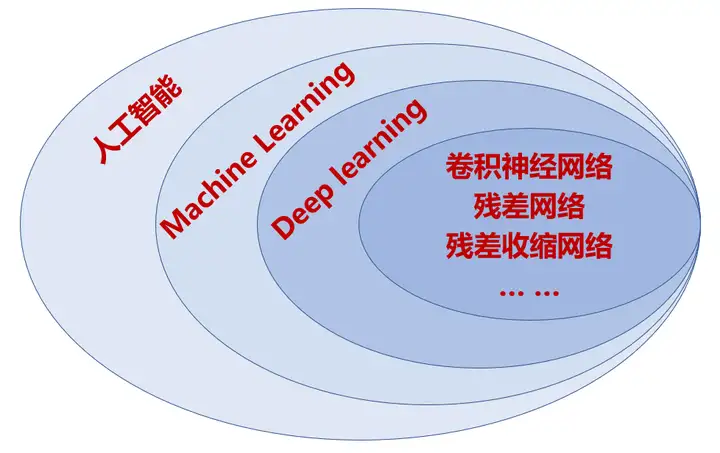
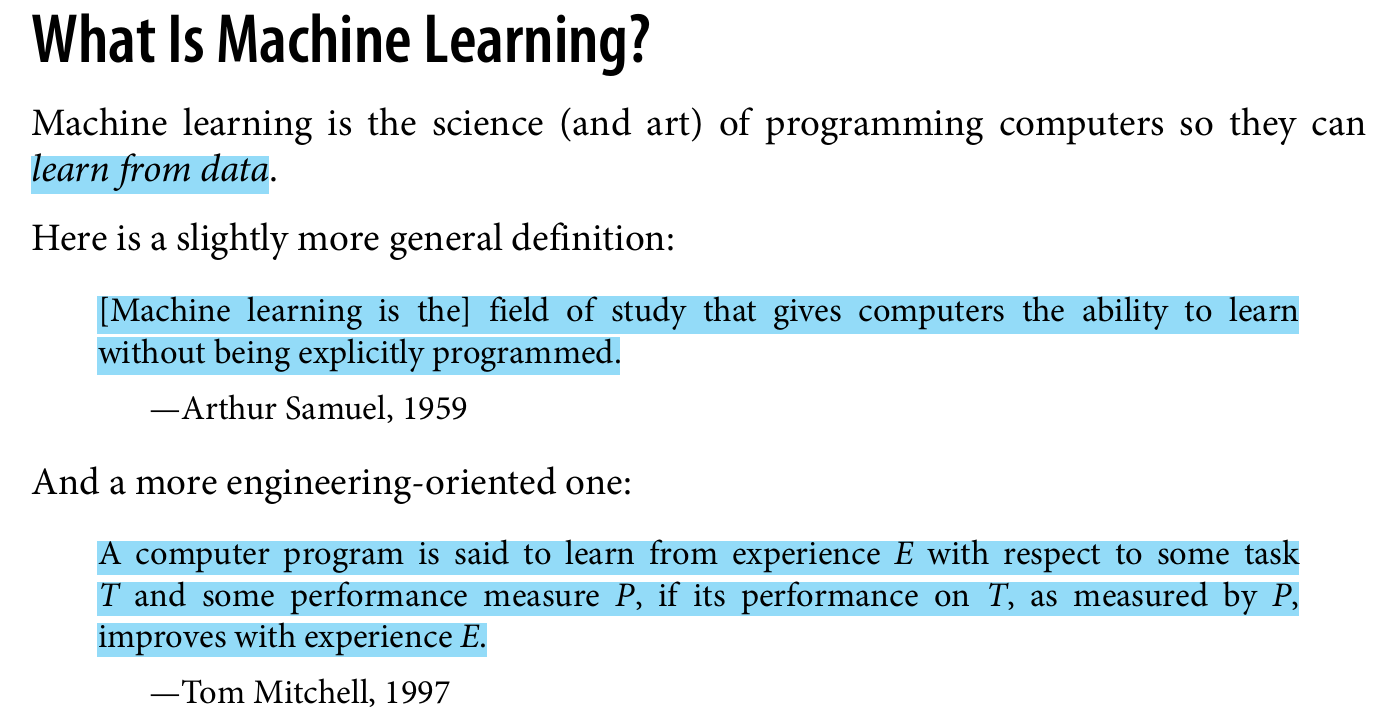
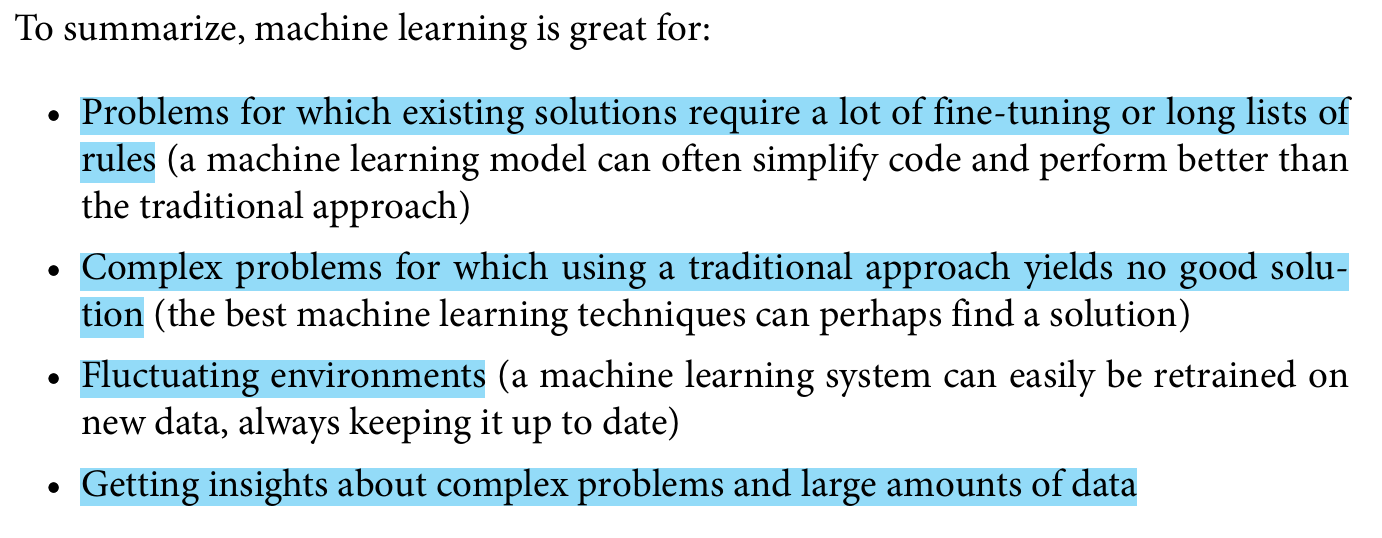
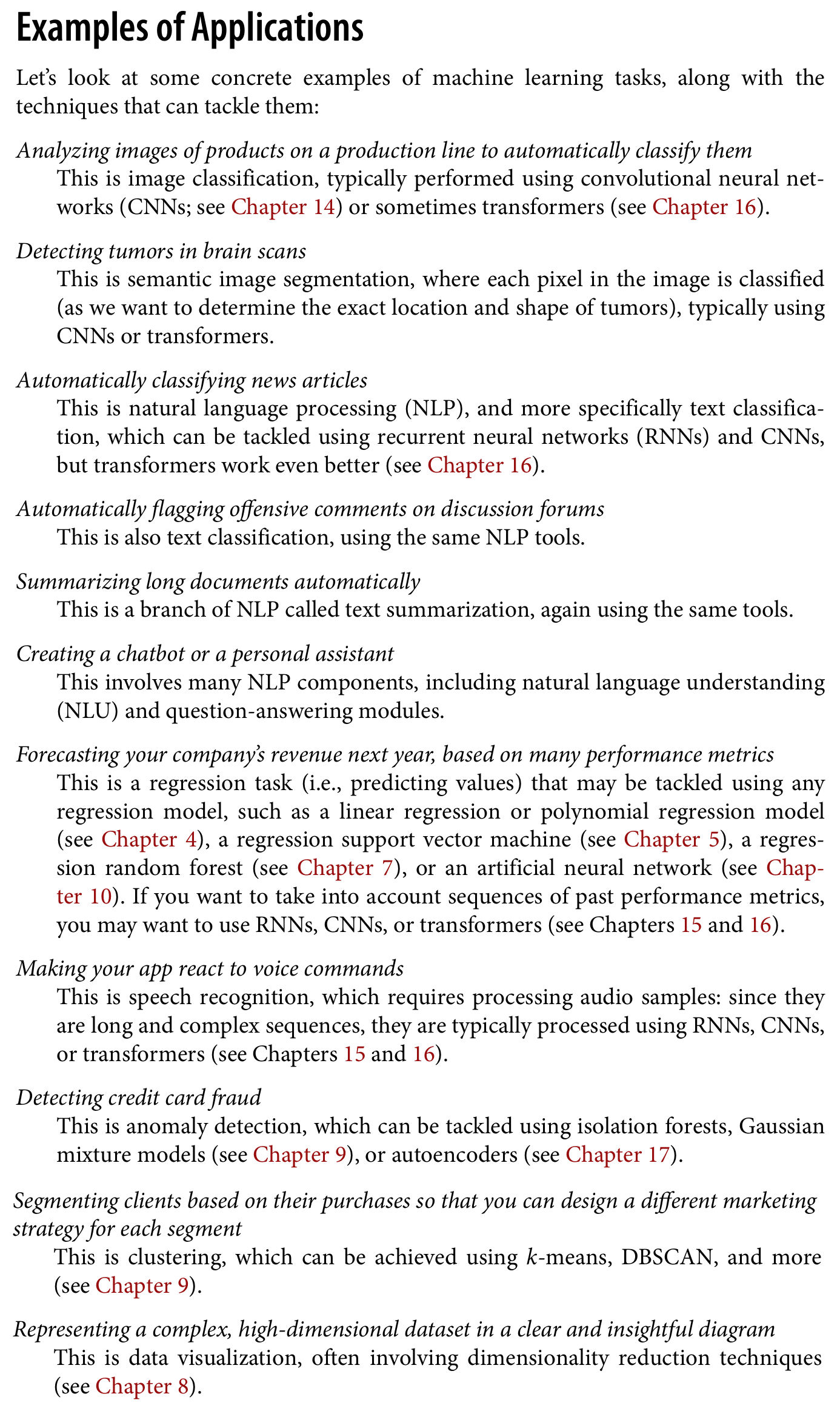


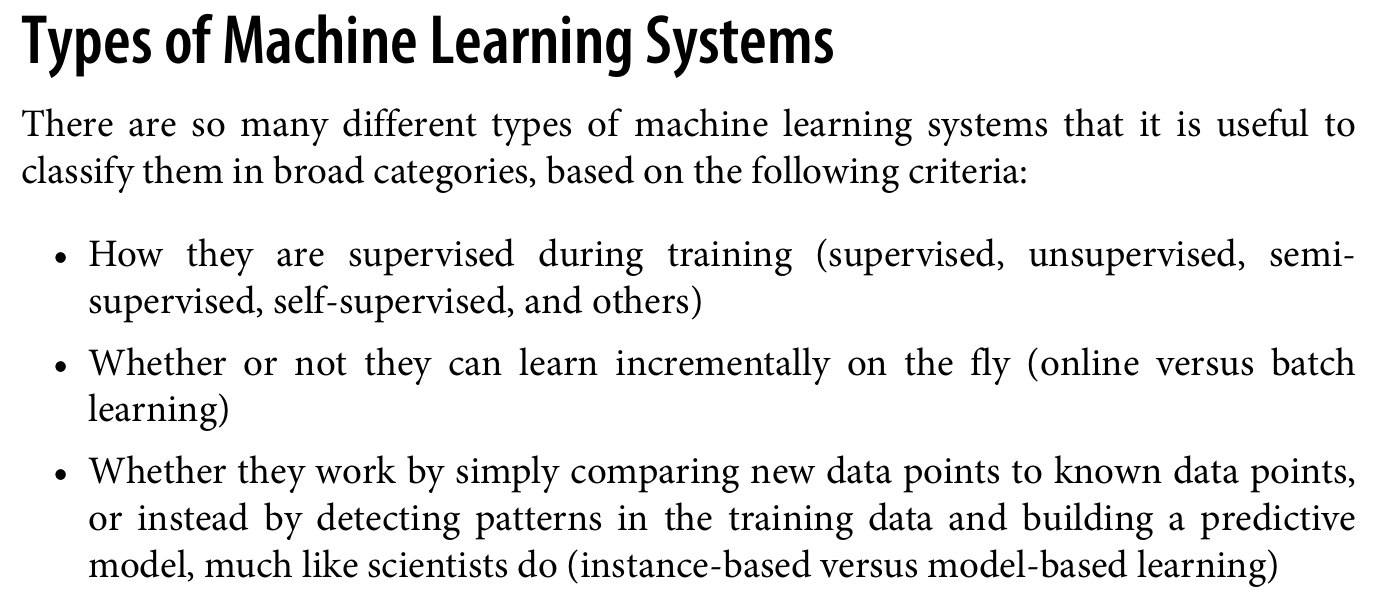
<https://www.bilibili.com/video/BV1owrpYKEtP/?spm_id_from=333.337.search-card.all.click&vd_source=44c160c96103b93e650b1a081c54ad17>

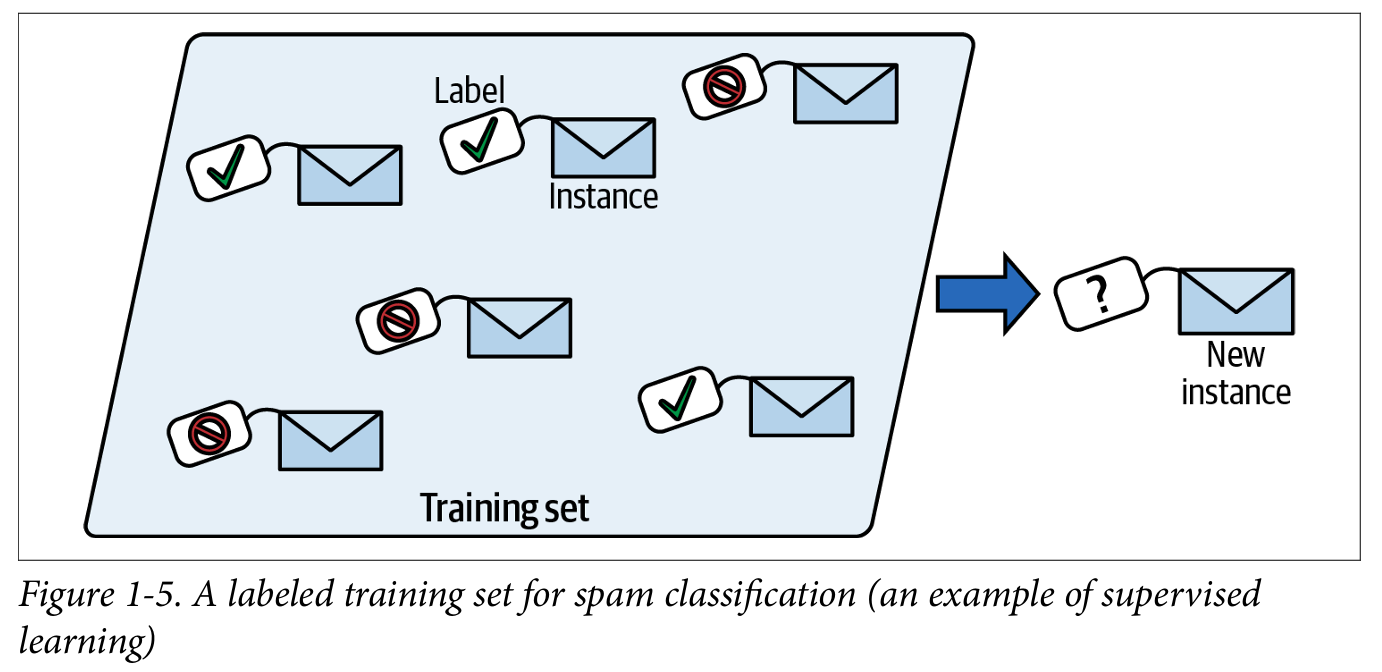


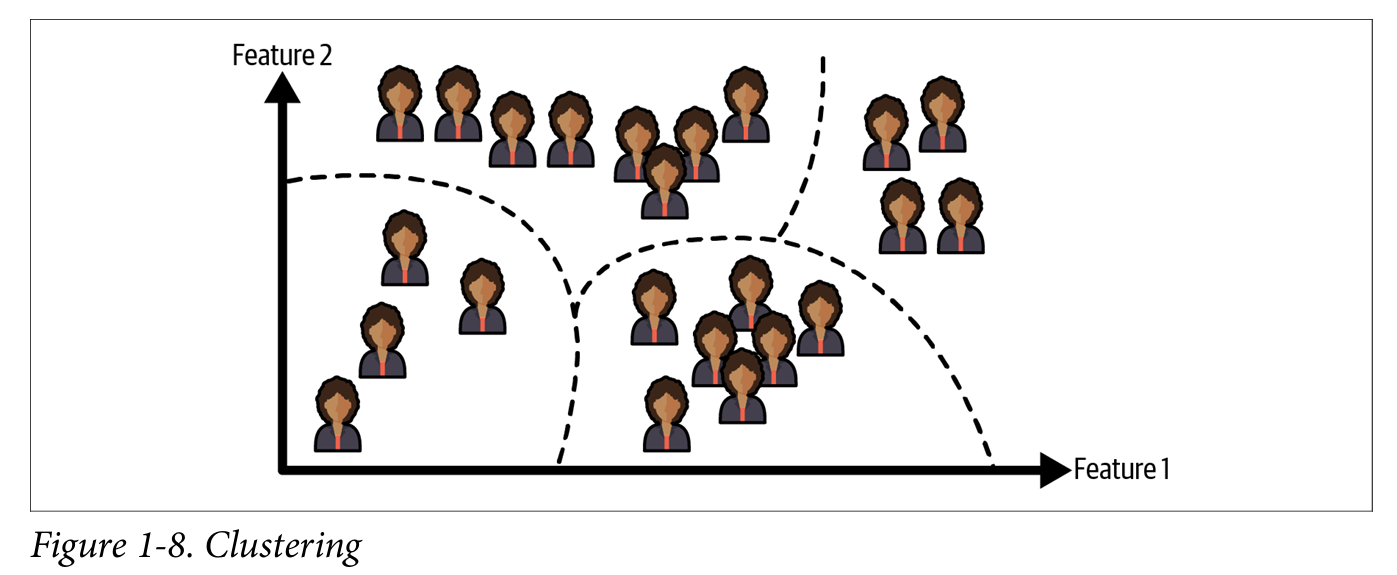
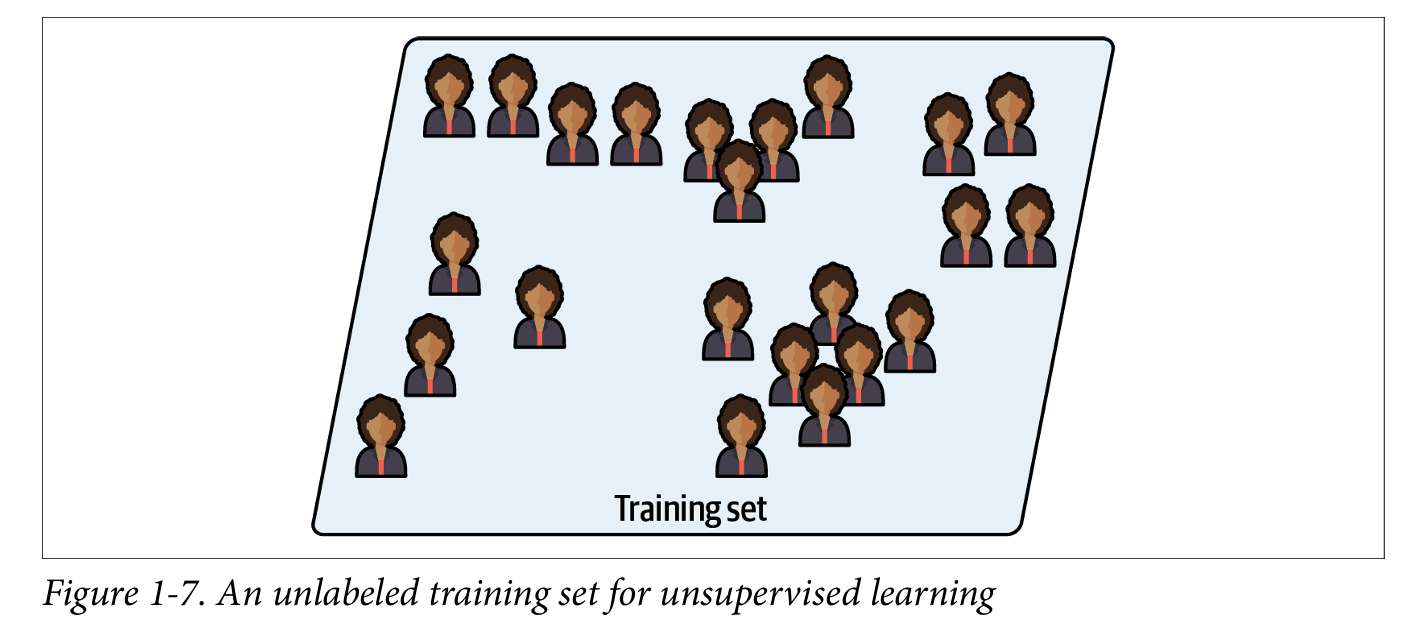


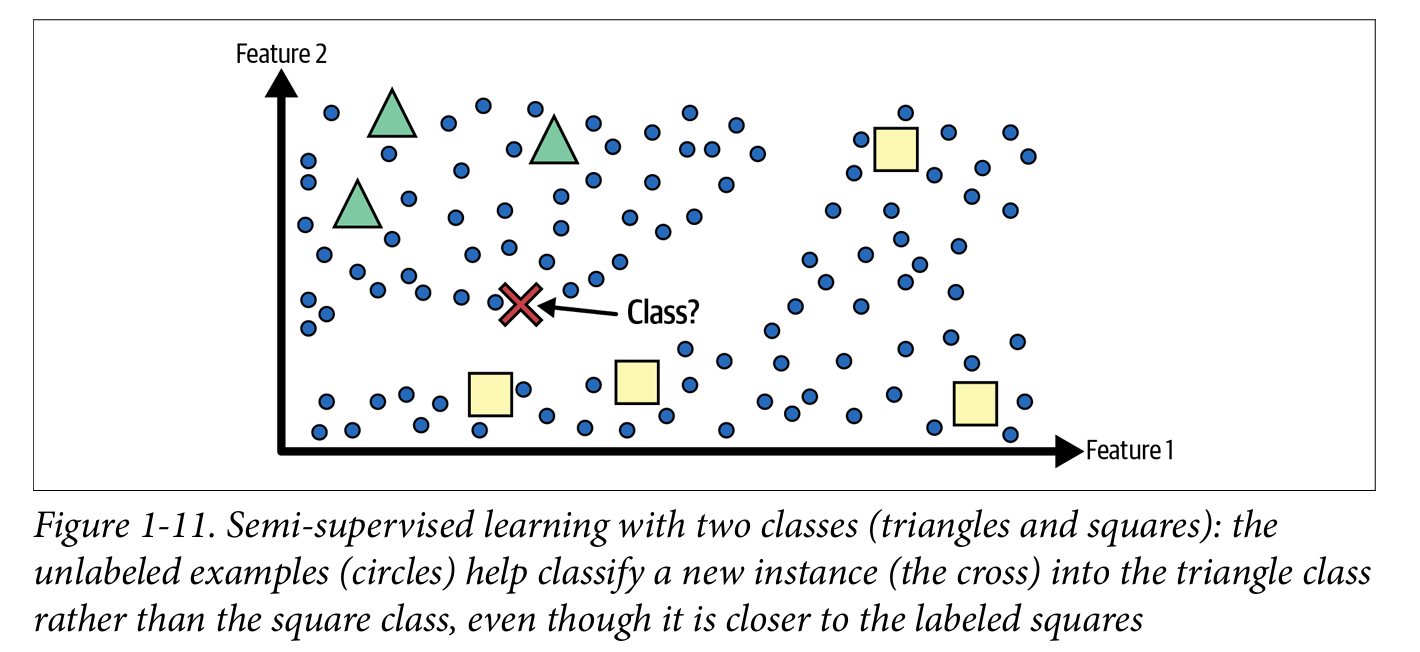


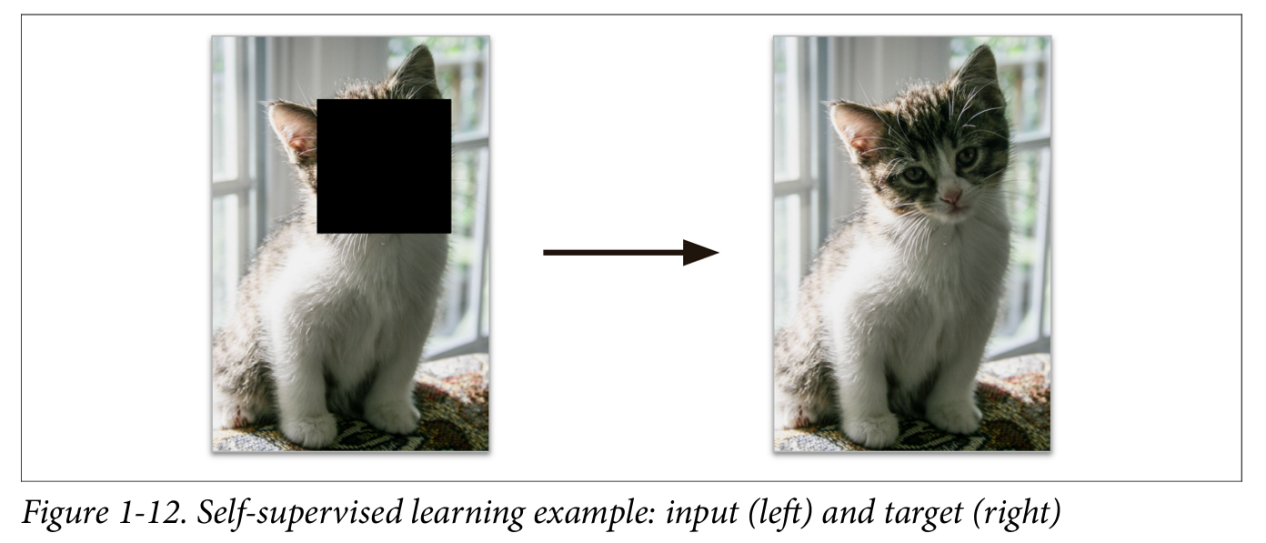


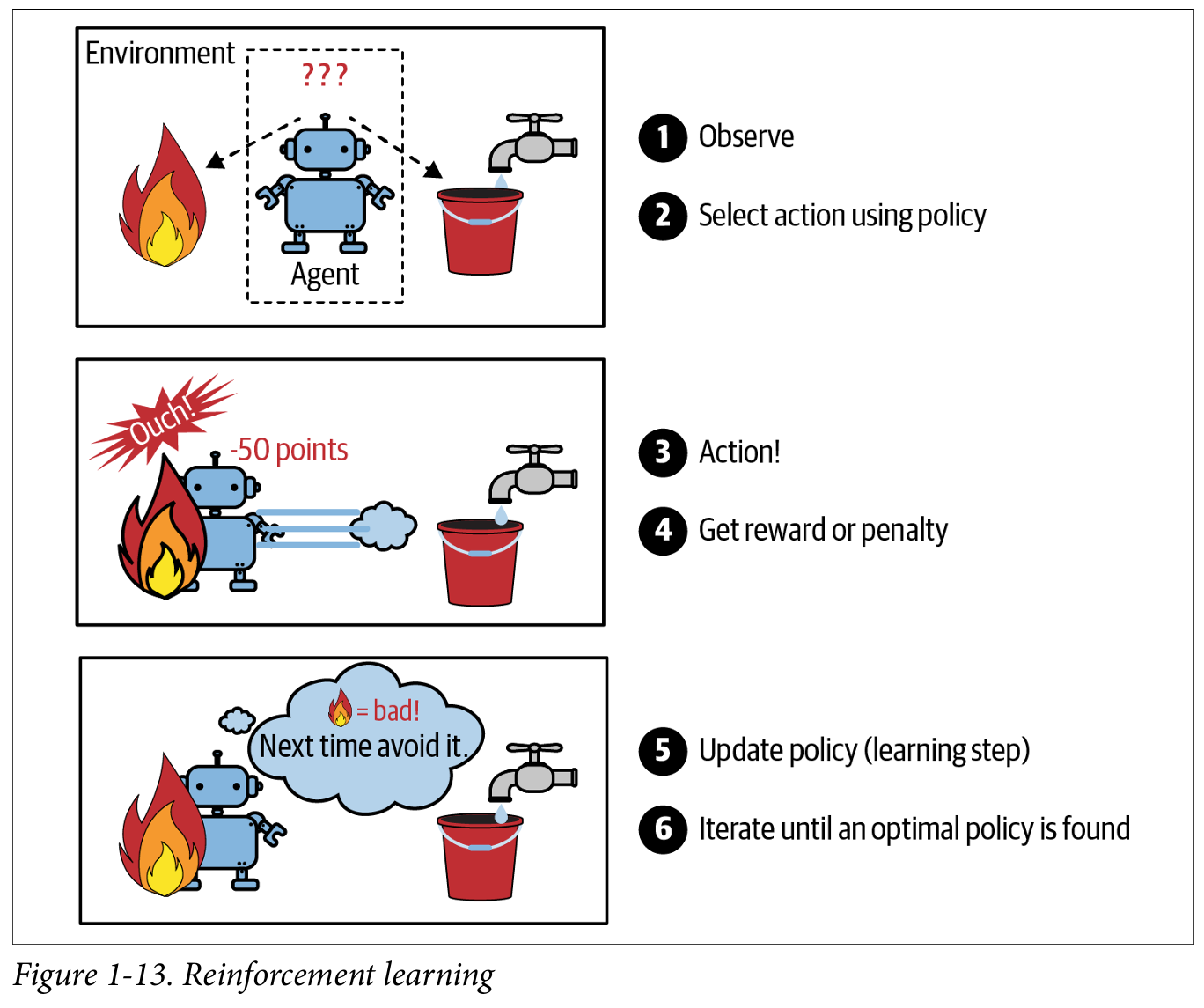


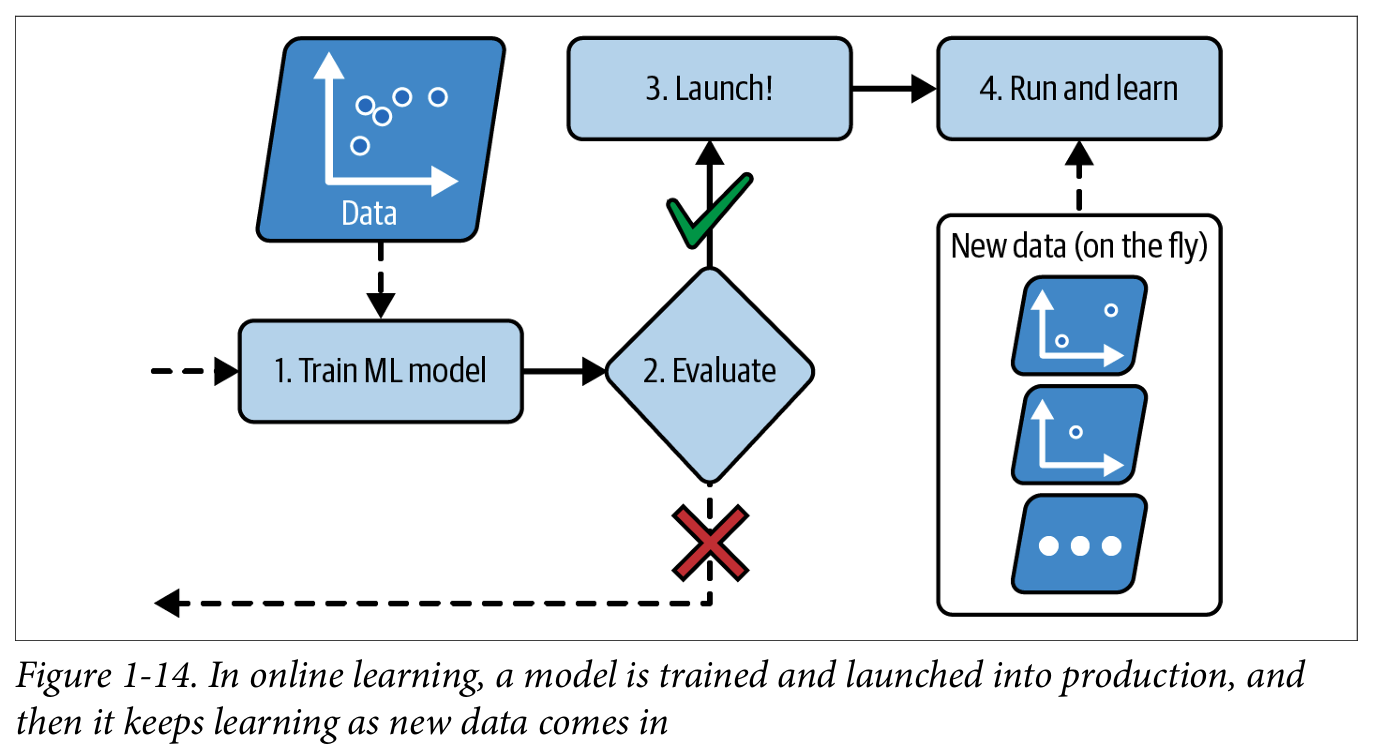


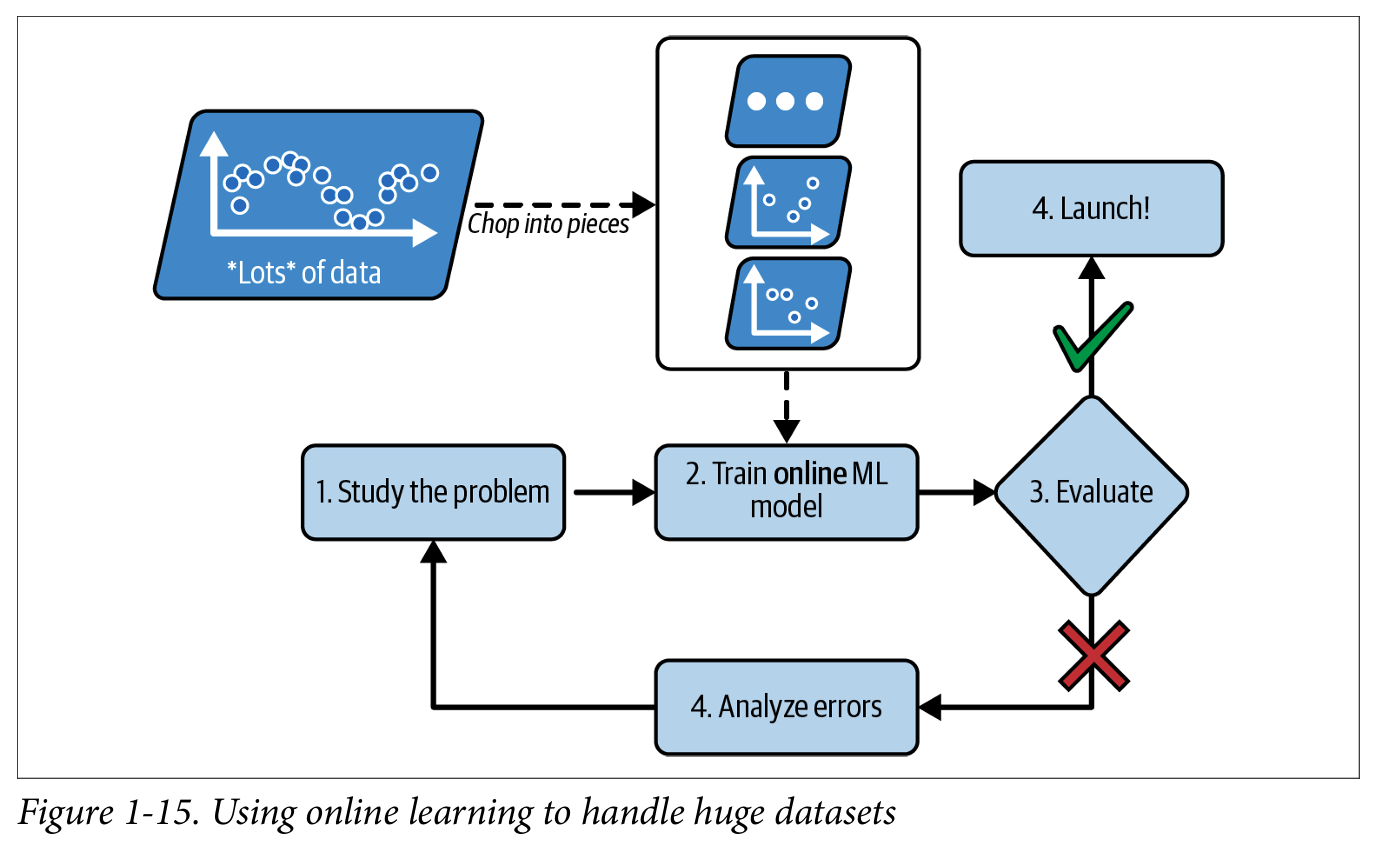


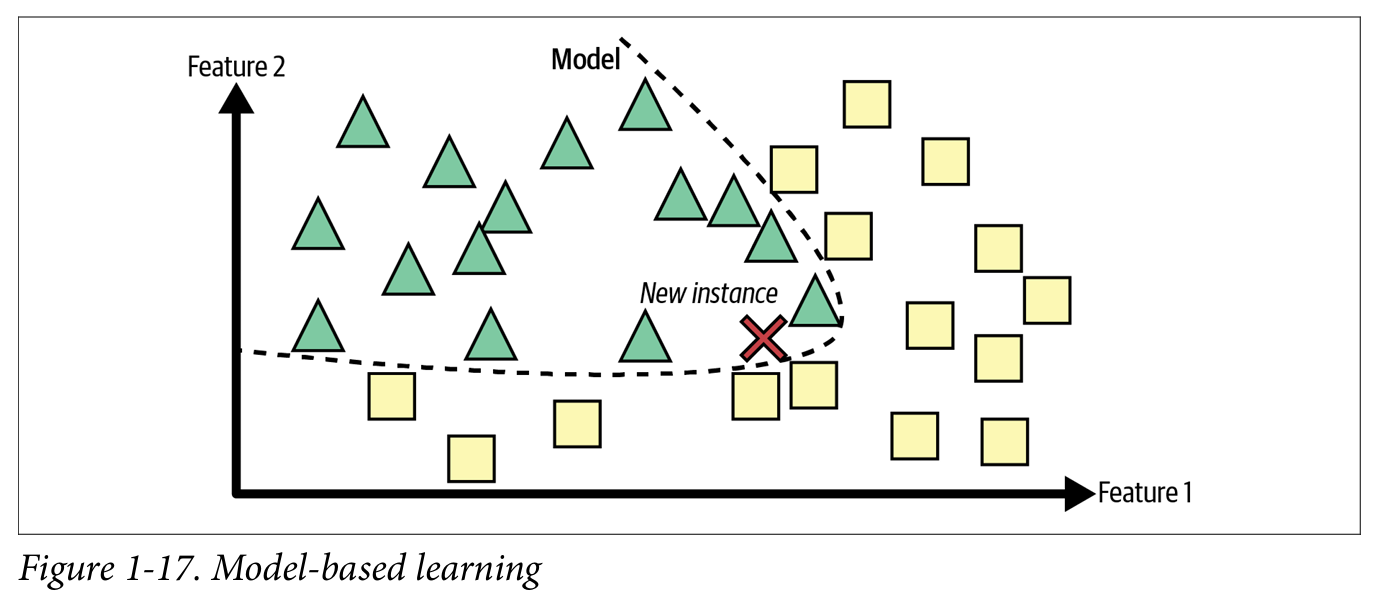
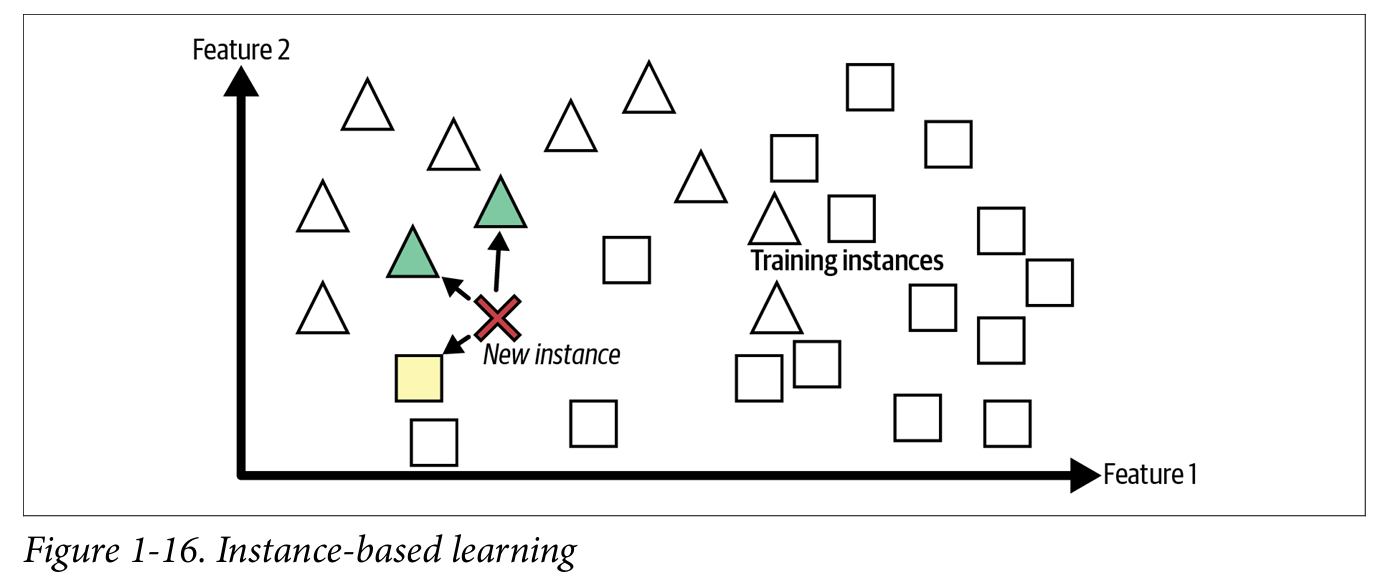


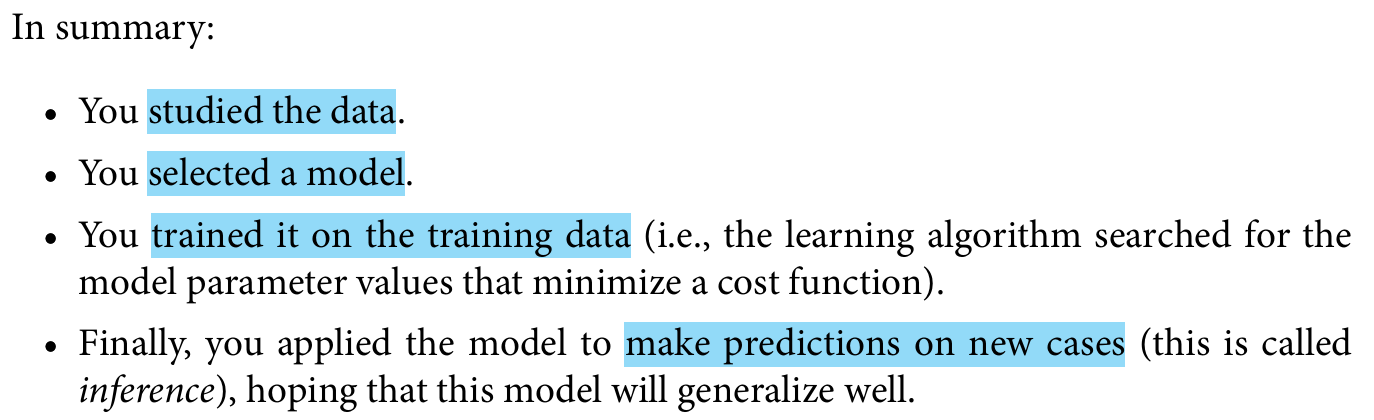


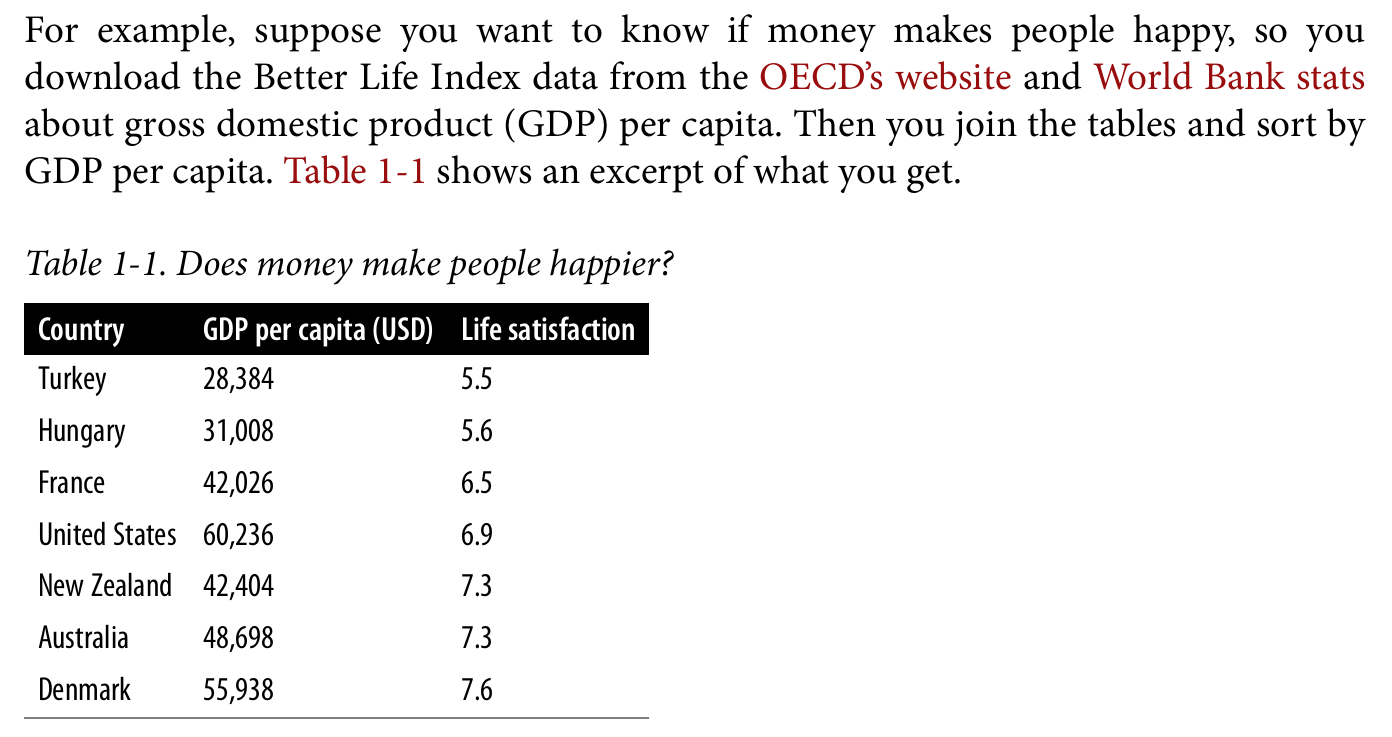


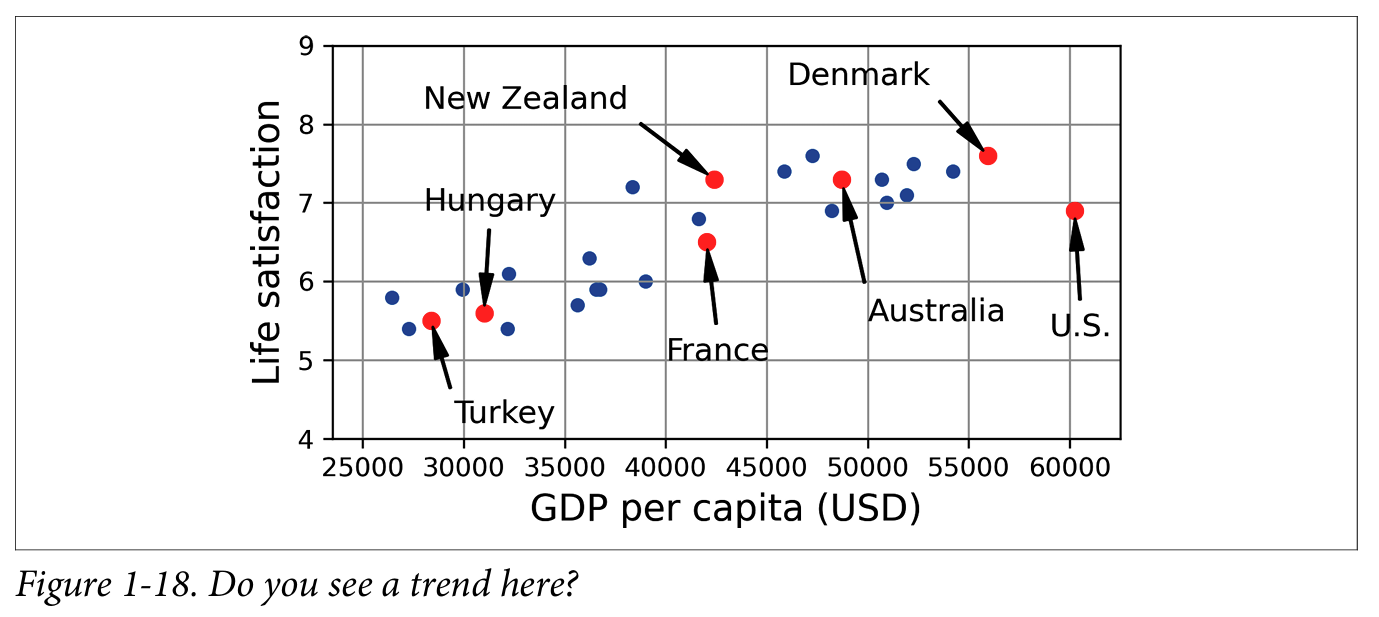


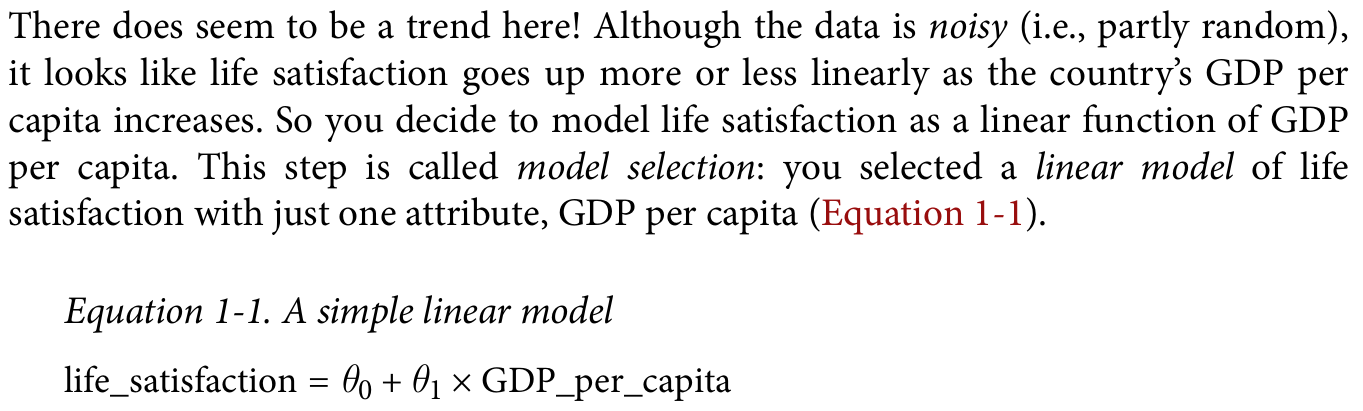


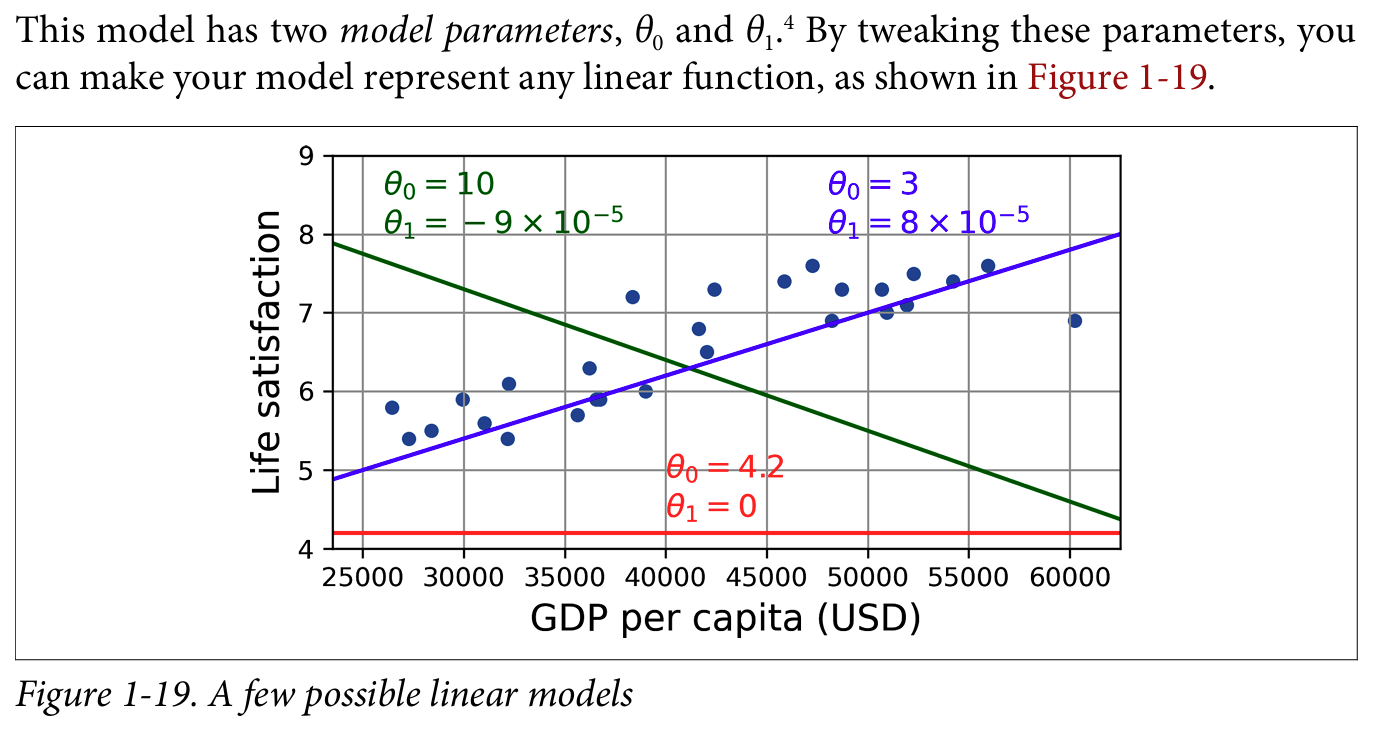


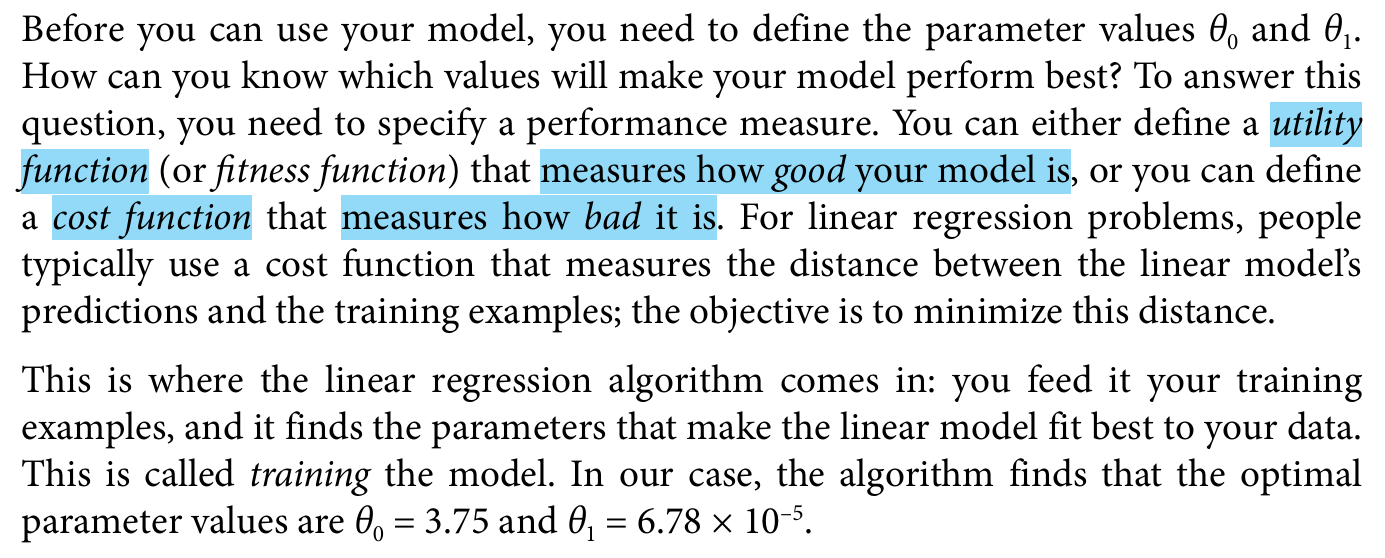


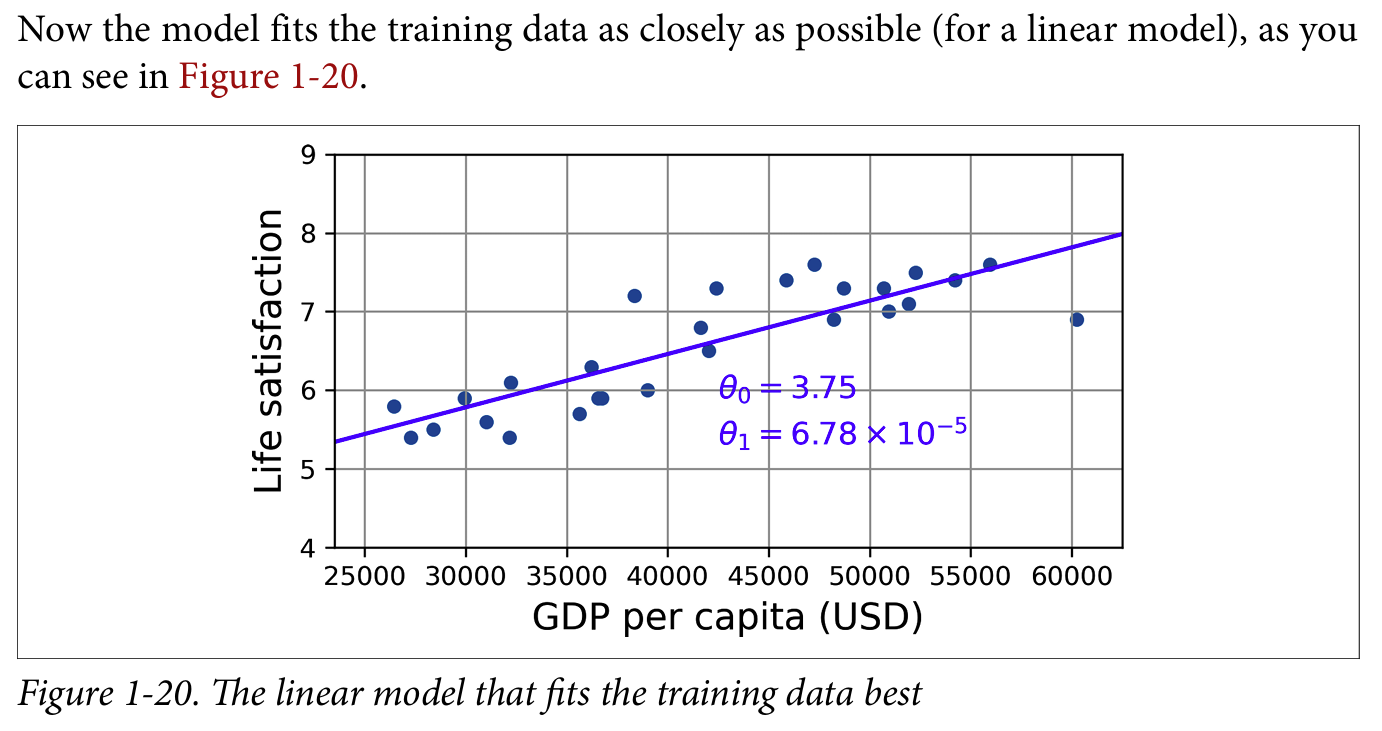


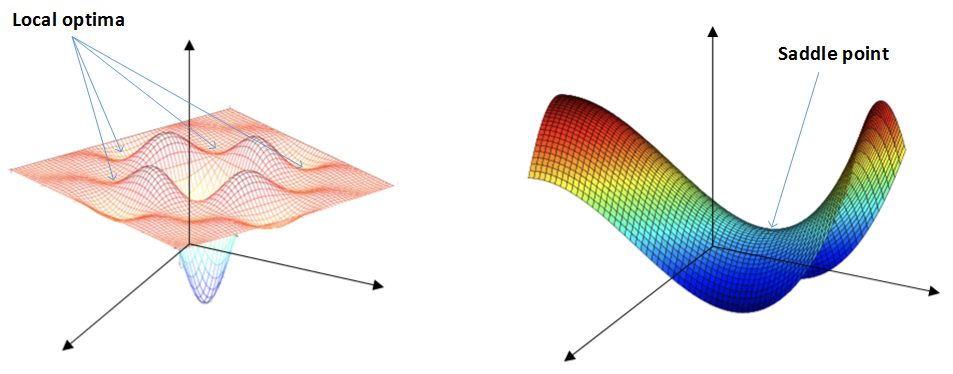




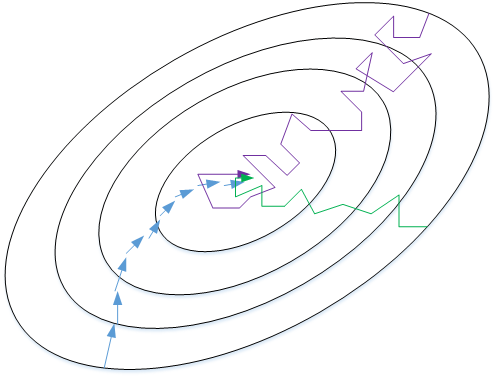








蓝色的线代表Batch gradient descent，紫色的线代表Stachastic gradient descent。Batch gradient descent会比较平稳地接近全局最小值，但是因为使用了所有m个样本，每次前进的速度有些慢。Stachastic gradient descent每次前进速度很快，但是路线曲折，有较大的振荡，最终会在最小值附近来回波动，难以真正达到最小值处。实际使用中，mini-batch size不能设置得太大（Batch gradient descent），也不能设置得太小（Stachastic gradient descent）。这样，相当于结合了Batch gradient descent和Stachastic gradient descent各自的优点，既能使用向量化优化算法，又能叫快速地找到最小值。mini-batch gradient descent的梯度下降曲线如下图绿色所示，每次前进速度较快，且振荡较小，基本能接近全局最小值。



一般来说，如果总体样本数量m不太大时，例如m≤2000，建议直接使用Batch gradient descent。如果总体样本数量m很大时，建议将样本分成许多mini-batches。推荐常用的mini-batch size为64,128,256,512。