

## 1.特征比数据量还大，选择什么分类器

线性分类器，因为维度高的时候，数据一般在维度空间里面会比较稀疏，很有可能线性划分

## 2.对于维度很高的特征，你是选择线性还是非线性分类器

理由同上

## 3.对于维度很低的特征，你是选择线性还是非线性分类器

非线性分类器，因为低维空间可能很多特征都跑到一起了，导致线性不可分

下面是吴恩达的见解：

1. 如果Feature的数量很大，跟样本数量差不多，这时候选用LR或者是Linear Kernel的SVM
2. 如果Feature的数量比较小，样本数量一般，不算大也不算小，选用SVM+Gaussian Kernel
3. 如果Feature的数量比较小，而样本数量很多，需要手工添加一些feature变成第一种情况