**《深度学习》花书学习**

# Week03总结

## 主要知识点



图1 花书前五章知识点

## 重点知识

虽然前五章有较多的知识点，但是这里主要总结有公式推导的部分，基础概念可以参照之前更新的消息。

### 线性回归：

任务：建立一个系统，根据输入可以预测输出；

经验：输入和真实输出。

性能度量1：均方误差（经验误差）；



正规方程

（normal equation）

性能度量2：结构误差；

（最小范数解）

对其进行求导



性能度量3：最大似然；



我们可以看出最大化关于的对数似然和最小化均方误差会得到相同的参数估计。但是对于相同的最优，这两个准则有着不同的值。这验证了MSE可以用于最大似然估计。

### 逻辑回归





为了表示方便，将权值向量和输入向量进行扩充，，则有



可以应用极大似然估计法估计模型参数，从而得到逻辑回归模型。

似然函数为



对数似然函数为



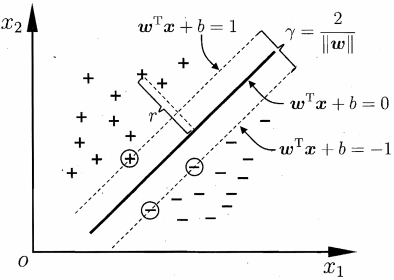
对其进行求极大值，得到的估计值。



所以有如下梯度下降法的迭代公式：



### 支持向量机





这就是支持向量机(Support Vector Machine，简称SVM)的基本型。

拉格朗日函数为：



求导得



将上式代回到对偶问题中（视频中因为没有用向量表示，所以不太好推导）



所以原问题的对偶问题如下：

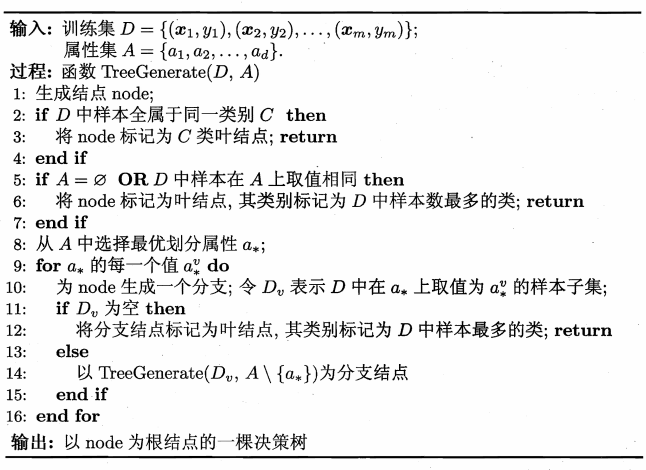
 （原问题的对偶）

KKT条件为。

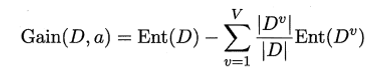
解出后，求出和即可得到模型



### 决策树

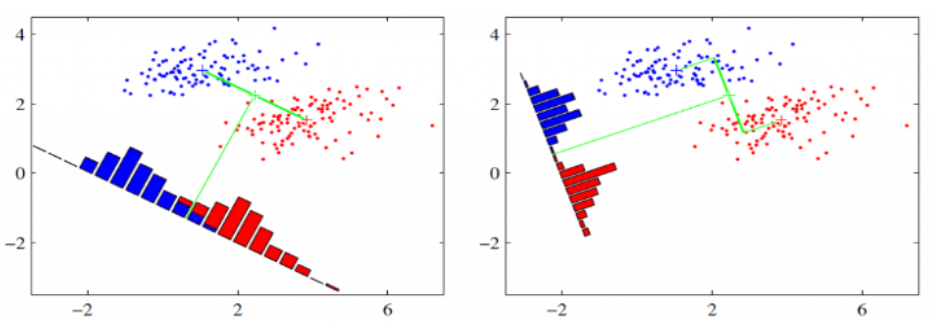


其中属性划分根据以下信息增益计算：



### 线性判别分析

线性判别式分析（Linear discriminant analysis），又称为Fisher线性判别。



通过调整权重向量组件，可选择一个投影方向，最大化分离类别。

设有K类样本，每一类的样本个数为，即对应第一类，对应第K类。

设为变换后的样本，则根据[PCA](#_PCA原理及推导)部分投影知识，可知



因此第K类之间的样本方差为：



均值

总体样本方差如下：



样本间的均值距离为：



则所有样本之间的均值距离为：



为了使得样本内的方差最小，且样本间的均值距离越大，则有下面的优化函数：



现假设=1，则等价于



构造拉格朗日函数：



求导得



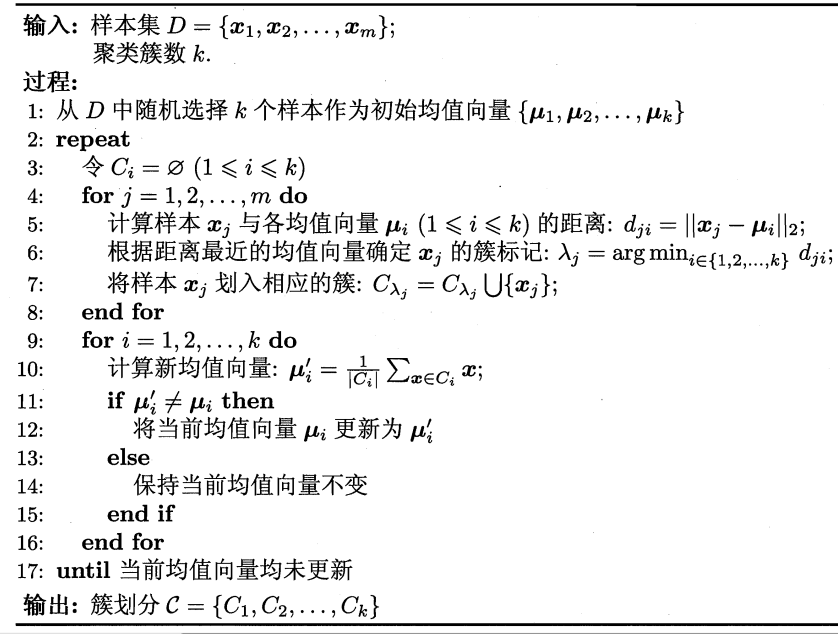
从而变成了对求特征值。

### k-均值

给定样本集，"k均值"(k-means)算法针对聚类所得簇划分最小化平方误差



其中是簇的均值向量。



### PCA

主成分分析是一种数据压缩或降维算法，其基本原理是确定新的方向向量（Vector direction），当投影所有的数据至该向量上时，希望投射平均均方误差尽可能地小（优化目标）。

如图2所示，A点需要x，y两个坐标来表示（二维平面空间），假设A在向量u上面的投影点为A’（垂直于向量u），则A’仅仅需要一个参数就能表示，就是OA’的长度（即A’在u上的坐标，因为此时u相当于一个新的坐标轴，A’在另一个垂直坐标轴上的坐标永远为0），我们就想着用A’来替换A，这样N个点（原来要2\*N个参数，对于M维空间就是M\*N），现在只需要（N+2）个参数（u也需要2个坐标参数，更严格地说只需要方向参数，但无所谓了，因为不管N+1还是N+2，都比原来降维了），但是此时就带来了误差，如AA’和BB’，所以我们要能够找到这样一个方向u，使得所有原始点与投影点之间的误差最小。

一般在PCA之前需要进行样本中心化，即每一维的数据都减去该维的均值。这样变换后每一维的均值都为0，这样相比于原始点，可以缩小与投影点的误差。



图2 投射原理

最小重构误差确定u的方向：

表示的投影向量，为单位向量，根据图3，要计算向量a在e方向的投影向量，首先计算a在e向上投影长度，根据余弦定理有，然后再乘以该方向上的单位向量即可。

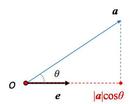


图3 投影计算

因此图5中，有



在n维空间中，x实际上就是n为列向量，点积操作可以写成矩阵相乘的形式（如图4所示），同时为表示简便，后面将省略“”，则有



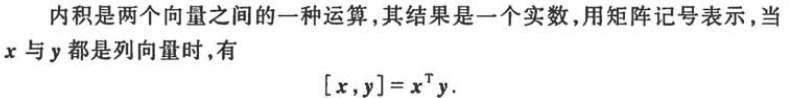
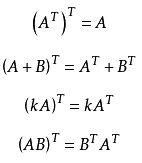


图4 点积定义

图5中的重构误差：

为实数，所以转置为本身，同时可以前提。



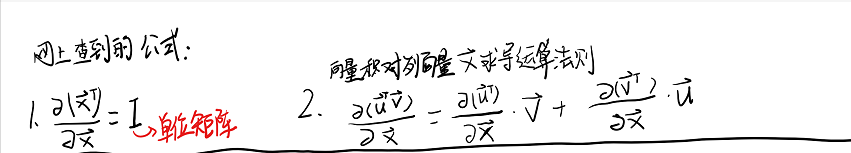
图5 误差计算推导

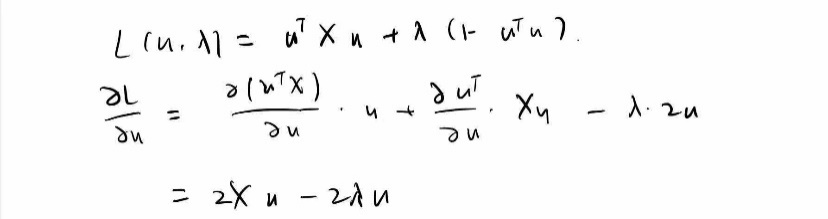
因此最小化重构误差（单样本）就等价于最大化，有



对于N个样本，则有



另，则

通过拉格朗日进行优化：





可以看出为的特征向量，为的特征值。