

Chapter 8 Ensemble Learning

Jinghao.Zhao

Abstract

This is the notes for ensemble learning, section 8.1 to 8.5, including formula and key points.

1 PCA

$$X = \begin{pmatrix} a_1 & a_2 & \dots & a_m \\ b_1 & b_2 & \dots & b_m \end{pmatrix} \quad (1)$$

$$[\frac{1}{m}XX^T = \begin{pmatrix} \frac{1}{m}\sum_{i=1}^m a_i^2 & \frac{1}{m}\sum_{i=1}^m a_i b_i \\ \frac{1}{m}\sum_{i=1}^m a_i b_i & \frac{1}{m}\sum_{i=1}^m b_i^2 \end{pmatrix}] \quad (2)$$

$$\begin{aligned} D &= \frac{1}{m}YY^T \\ &= \frac{1}{m}(PX)(PX)^T \\ &= \frac{1}{m}PXX^TP^T \\ &= P(\frac{1}{m}XX^T)P^T \\ &= PCP^T \end{aligned} \quad (3)$$

这里再解释一下为什么最后就能通过特征分解求得最优解。在之前我们需要先了解一下拉格朗日乘法。根据拉格朗日乘法,

$$\min -tr(W^TXX^TW)$$

$$s.t. W^TW = 1$$

上述问题等价于:

$$\min W^TXX^TW$$

$$s.t. W^TW = 1$$

可以得到拉格朗日函数:

$$F(W, \lambda) = -W^TXX^TW + \lambda W^TW$$

对W求偏导数有：

$$\begin{aligned}\frac{F(W, \lambda)}{\partial W} &= \frac{-W^T X X^T W + \lambda W^T W}{\partial W} = 0 \\ -2X X^T W + 2\lambda W &= 0 \\ X X^T W &= \lambda W\end{aligned}$$

最后的这个式子正是特征值分解的式子！（ $Av = \lambda v$ ）所以我们只需要对协方差矩阵 XX^T 进行特征值分解（非方阵的时候使用奇异值分解），将求得特征值排序，再去前k个特征值对应的特征向量，组成矩阵W，这就是我们PCA最终得到的解。