说话人识别

1 UBM-ivector （感觉就是对每段语音进行降维，得到一个embedding向量，这一部分也可以用网络训练）



m: m是一个通用的矢量。训练UBM通用背景模型，通常我们会先训练一个协方差对角矩阵，再训练全矩阵。**UBM特点**：一个与说话人、信道无关的高斯混合模型；可以作为这个训练集的统一参考坐标空间；

M: 待提取的语音的高斯均值超矢量

T:是一个变化矩阵，类似降维

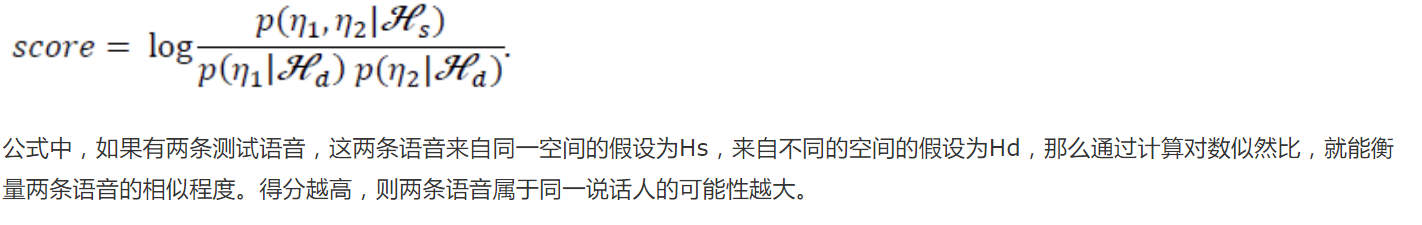
w： 是抽取的ivector特征

2. 分类器（测试时，一般要是一个1对N的问题，先用少量特定说话人的数据得到每个人的向量表示，然后对新来的样本进行测试分类性能。）

PLDA （要训练去获得参数，这个其实是简化版本，没有对噪声建模了）

PLDA 公式 : Ivector = 数据均值+ 本征音空间\*说话人因子+epsilon

对每个说话人的ivector，建模，降维。然后对说话人因子进行分类。然后用对数丝然函数log(音频1,音频2| 是同一个人)/P(音频1|甲)\*P(音频2|乙)的分数来进行判别和分类哦。（分类时，有Hs Hd, 同一空间假设和不同空间假设）



LDA （降维，得到一个与距离有关的隐藏表示，也是计算余弦距离）

Cosine distance

先用少量特定说话人的数据得到每个人的向量表示（例如对ivector取平均），然后对新来的样本进行测试分类性能（例如计算余弦距离）