Fisherfaces recognition

周强

2017 年 7 月 20 日

计算机视觉、模式识别和机器学习中的一个关键问题是为手头的任务定义适当的数据表示方式。

表示输入数据的一种方式是找到代表大部分数据方差的子空间。这可以通过使用主成分分析（PCA）获得。当应用于脸部图像时，PCA 产生一组特征脸（eigenfaces）。这些特征脸是与训练数据的协方差矩阵的最大特征值相关联的特征向量。因此找到的特征向量对应于最小二乘法（LS）解。这确实是表示数据的有效方法，因为它确保维持数据差异，同时消除样本向量中的原始特征（维度）之间不必要的现有相关性。

当目标是分类而不是表示时，LS 解决方案可能不会产生最理想的结果。在这种情况下，希望找到一个子空间，该子空间将相同类别的样本向量映射到特征表示中的单个点，并且不同类别的样本向量尽可能远离彼此。为实现这一目标而衍生的技术被称为判别分析（DA）。

最着名的 DA 是线性判别分析（LDA），可以从 R.A.Fisher 在 1936 年建议的想法中得出。当 LDA 用于找到一组面部图像的子空间表示时，所得到的定义该空间的基向量（basis vectors）称为Fisherfaces。

图 1: 这里显示的是来自一组 100 个类（目标）的前四个 Fisherfaces

1 Discriminant Scores

为了计算 Fisherfaces，我们假设每个类中的数据是正态分布的。我们将多元正态分布表示为 $N_i(\mu_i \ \sum_i)$，其中 μ_i 为均值，$\sum_i$ 为协方差矩阵，$f_i(x|\mu_i, \sum_i)$ 为概率密度函数。

在 C 类问题中，我们有 $N_i(\mu_i \ \sum_i)$，其中 $i = 1 \ \cdots \ C$。给定这些正态分布及其类别先验概率 P_i，通过比较所有的对数似然 $f_i(x|\mu_i, \sum_i)P_i$ 来给出测试样本 x 的分类。也就是，

$$\arg \min_{1 \leq i \leq C} d_i(x) \tag{1}$$

其中 $d_i(x) = (x - \mu_i)^T \sum_i^{-1}(x - \mu_i) + \ln|\sum_i| - 2\ln P_i$ 被称为每个类的判别得分。由此定义的判别得分产生贝叶斯最优解。

判别得分通常导致类之间的二次分类边界。然而，对于所有协方差矩阵相同的情况，$\sum_i = \sum, \forall i$，d_i 的二次部分消失，产生线性分类器。这些分类器称为线性判别基，因此，线性判别分析的名称由来。所有协方差相同的情况称为同方差正态分布。

假设 C = 2，并且类是同方差正态分布。将样本特征向量投影到与判别得分给出的分类超平面正交的一维子空间上。结果是，在 p 维的原始空间中和在仅一维的该子空间中的错误分类样本的数量是相同的。这很容易验证。由于分类边界是线性的，所以在空间一侧的所有样本将保留在 1 维子空间的同一侧。这个重点是 R.A. Fisher 首次提出的，从而允许我们定义 LDA 算法和 Fisherfaces。

2 计算 Fisherfaces

上一节中给出的理论论证显示了如何获得 2-class 同方差的贝叶斯最优解。一般来说，我们将有两个以上的类。在这种情况下，我们将上述问题重新定义为尽量减少类内（within-class）差异并最大限度地提高类间（between-class）距离。

类内差异可以使用类内分散矩阵 (scatter matrix) 来估计

$$S_w = \sum_{j=1}^{C}\sum_{i=1}^{n_j}(x_{ij} - \mu_j)(x_{ij} - \mu_j)^T \tag{2}$$

其中 x_{ij} 是第 j 类的第 i 个样本，μ_j 是类 j 的平均值，n_j 是类 j 中的样本数。

类似地，类间差异是使用类间分散矩阵计算的

$$S_b = \sum_{j=1}^{C} N_j(\mu_j - \mu)(\mu_j - \mu)^T \tag{3}$$

其中 μ 表示所有类的平均值。

我们现在想找到那些基本向量 V，其中 S_w 被最小化并且 S_b 被最大化。矩阵 V 的列矢量 v_i 是定义子空间的基本向量。可以如下计算：

$$V_{opt} = \arg\max_V \frac{|V^T S_b V|}{|V^T S_w V|} \tag{4}$$

该问题的解决方案由广义特征值分解给出

$$S_b V = S_w V \Lambda \tag{5}$$

其中 V 是（如上所述）特征向量的矩阵，Λ 是相应特征值的对角矩阵。

与非零特征值相关联的 V 的特征向量是Fisherfaces。最多有 C-1 个 Fisherfaces。这可以从 S_b 的定义中容易地看出。注意，在我们的定义中，S_b 是 C 个特征向量的组合。任何 C 个向量定义 C-1 或更小尺寸的子空间。当这些向量彼此线性相互独立时，该等式成立。