实验一 手写数字概率模型参数估计与识别

2019211039 卢建中

一、 实验目的

学会对实际数据建模,并利用最大似然方法估计分布的参数。建模完成后结合贝叶斯准则完成分类识别任务，并对所建模型分析评判。

二、 实验内容

1、对各类图片的分布进行建模,在训练集上利用最大似然方法,估计其分布参数;

2、利用贝叶斯判决对测试集上每张图片进行分类,并验证分类准确率;

3、对实验结果进行分析。

三、 实验原理

1、 PCA 主成分分析

主成分分析是一种数据分析方法，其出发点是从一组特征中计算出一组按重要性从大到小排列的新特征，它们是原有特征的线性组合，并且相互之间是不相关的。

记$x_1,\dots,x_p$为p个原始特征，设新特征$\xi_i, i=1,\dots,p$是这些原始特征的线性组合：

$$\xi_i=\sum_{j=1}^{p}\alpha_{ij}x_j=\vec{\alpha}_i^T\vec{x}$$

为了统一尺度，不妨要求线性组合系数的模为 1，即：

$$\vec{\alpha}_i^T\vec{\alpha}_i=1$$

则可以得到：

$$\vec{\xi}=\vec{A}^T\vec{x}$$

其中，$\vec{\xi}$是由新特征ξ_i组成的向量，$\vec{A}$是特征变换矩阵。要求解的是最优的正交变换$\vec{A}$，它使新特征ξ_i的方差达到极值。正交变换保证了新特征之间不相关，而新特征的方差越大，则样本在该维特征上的差异就越大，因而这一特征就越重要。

考虑第一个新特征ξ_1：

$$\xi_i = \sum_{j=1}^{p} \alpha_{1j} x_j = \vec{\alpha}_1^T \vec{x}$$

它的方差是：

$$\mathrm{var}(\xi_1) = E[\xi_1^2] - E[\xi_1]^2 = \vec{\alpha}_1^T \Sigma \vec{\alpha}_1$$

其中，Σ是 x 的协方差矩阵，可以用样本来估计；要在系数模为 1 的条件下最大化方差，这等价于求下列拉格朗日函数的极值：

$$\mathrm{f}(\vec{\alpha}_1) = \vec{\alpha}_1^T \Sigma \vec{\alpha}_1 - v(\vec{\alpha}_i^T \vec{\alpha}_i - 1)$$

v 是拉格朗日乘子。将上式对$\vec{\alpha}_1$求导并令它等于零，就得到$\vec{\alpha}_1$的最优解满足：

$$\Sigma \vec{\alpha}_1 = v \vec{\alpha}_1$$

这是协方差的特征方程，即$\vec{\alpha}_1$一定是矩阵Σ的本征向量，v 是对应的本征值。代入上面的方程，可以得到方差的表达式：

$$\mathrm{var}(\xi_1) = \vec{\alpha}_1^T \Sigma \vec{\alpha}_1 = v$$

因此，最优的$\vec{\alpha}_1$应该是Σ的最大本征值对应的本征向量。ξ_1称作第一主成分，它在原始特征的所有线性组合里是方差最大的。

在后面确定后续的新特征时，除了需要满足以上的表达式，还需要和之前的主成分不相关，实际上就是对应于本征向量从大到小排列。

2、最大似然估计

我们把要估计的参数记为θ，它是确定但未知的量（多个参数时是向量）。每类样本集记作$\mathbf{x}_i, i = 1, \ldots, c$，其中的样本都是从密度为$\mathrm{p}(x|w_i)$的总体里中独立抽取出来的，即所谓满足独立同分布条件。类条件概率密度$\mathrm{p}(x|w_i)$具有某种确定的函数形式，只是其中的参数θ未知。假设样本包含 N 个样本：

$$\mathbf{x} = \{x_1, \ldots, x_N\}$$

由于样本是独立地从$\mathrm{p}(x|w_i)$中抽取的，所以在概率密度为$\mathrm{p}(x|w_i)$时获得的样本集的概率即出现$\mathbf{x}$中各个样本的联合概率密度是：

$$\mathrm{l}(\theta) = p(\mathbf{x}|\theta) = p(x_1, \ldots, x_N|\theta) = \prod_{i=1}^{N} p(x_i|\theta)$$

我们将上式称为未知参数的似然函数。

一般来说，使似然函数的值最大的$\hat{\theta}$是样本$x_1, \ldots, x_N$的函数，记为$\hat{\theta} = d(x_1, \ldots, x_N)$，它被称为$\theta$的最大似然估计。

正态分布下的最大似然估计：

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\hat{\Sigma} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{\mu})(x_i - \hat{\mu})^T$$

四、 实验结果及分析

1、首先对数据进行 pca 降维：

```
%% PCA降维
K = 10;  %维度
train = cell(10,1);
train_norm = data_central(train_set); %中心化数据

[U, S] = eig((1/size(train_norm,1)) * train_norm' * train_norm);  %特征值分解
train_norm_project = projectData(train_norm, U, K); %映射到相应维度
```

图 1 pca 降维

因为主成分分析是无监督学习，此时应该对原始数据直接全部进行降维，而不是先将对应数字的数据集分开，然后再降维。

2、假设投影后的数据按照高斯分布，根据训练集的标签进行高斯分布参数均值与协方差的的估计

```
%% 高斯分布参数估计
for i=1:10
    train{i,1} =  train_norm_project(find(train_label==i-1),:);
    mu(i,:) = mean_own(train{i,1}); %估计均值
    sigma(:,:,i) = (train{i,1}-mu(i,:))'*(train{i,1}-mu(i,:))/size(train{i,1},1);
    %估计协方差
end
```

图 2 高斯分布参数估计

3、利用贝叶斯公式进行判决

```
%% 贝叶斯决策

test_norm_project = projectData(test_set, U, K);

for i = 1:10 %先验概率
    predic(i) = size(train{i,1},1)/size(train_set,1);
end

for i = 1:size(test_label)
    for j = 1:10
    gauss_bayes(i,j) = -0.5*(test_norm_project(i,:)-mu(j,:))*inv(sigma(:,:,j))*...
        (test_norm_project(i,:)-mu(j,:))'+log(predic(1,j))-0.5*log(det(sigma(:,:,j)));
    %高斯分布
    [a,result(i)] = max(gauss_bayes(i,:));
    end
end
```

图 3 贝叶斯决策

在代入贝叶斯公式时，不要忘记计算先验概率。

4、准确率计算

```
right = 1-size(find(result'-test_label-1~=0),1)/size(test_label,1);
```

图 4 准确率计算

5、实验结果：

考虑降维维度与准确率的关系：

表一：准确率与维度

维度 K	5	10	20	30	40	50
准确率	0.7320	0.8918	0.9504	0.9609	0.9625	0.9636
维度 K	60	100	150	180	200	300
准确率	0.9612	0.9570	0.9507	0.1135	0.1135	0.0980

表二：计算时间与维度

维度 K	5	10	20	30	40	50
时间/s	4.93	5.49	6.43	8.21	9.53	12.72
维度 K	60	100	150	180	200	300
时间/s	15.3	39.54	77.1	112.02	149.3	312．83

注：运行环境为 matlab2018b，cpu:i7-6600u,内存 16g.

从上述表格中可以看到，随着降维维度的增加，准确率先上升，随后再下降，在 50 左右达到最大值。当降维较大时，则保留了较多的无效信息,包括噪声，因此不利于分类，其准确率必然会下降。而当所降维度较低时，该主成分忽略掉了其他的一些特征，正确率也不会很高。

随着降维的维度的增加，运算的时间也成倍的增加，因此合理选择 K 至关重要。

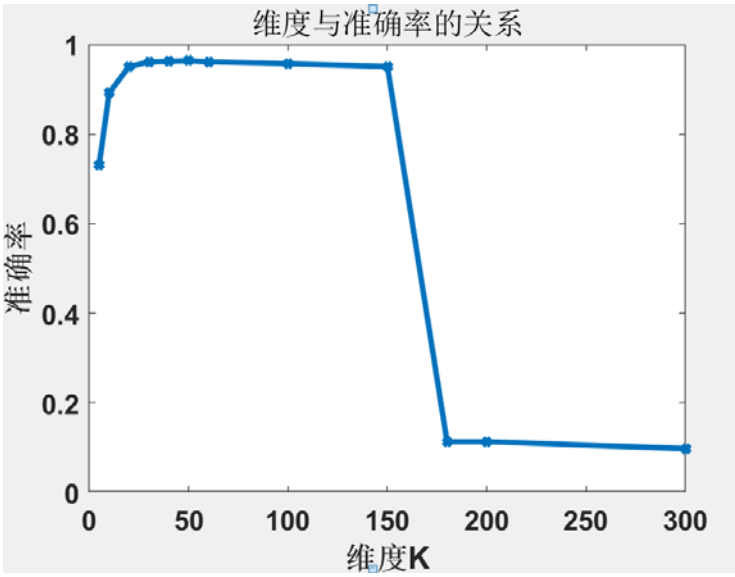

图 5 维度与准确度的关系

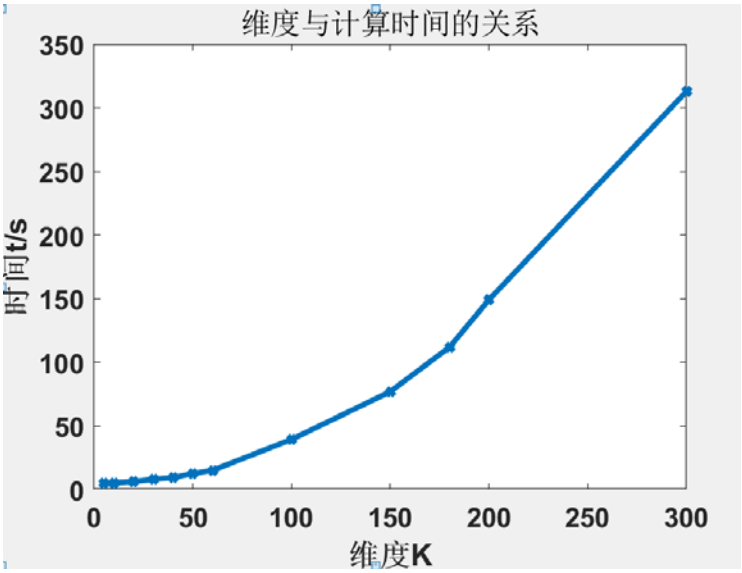

图 6 维度与计算时间的关系

6、实验改进

试想在实验进行贝叶斯概率计算的时候，可以将串行的循环计算尝试该进程并行的矩阵运算，可以有效加快效率。但是在后续的改进过程中，仍然有一些 bug 没有调好，还需要继续尝试。

在实验绘图时，可以尽量多取些维度，这样让曲线更加平滑而且相应的特征更加明显。

五、 实验总结

通过本次实验，加深了对于主成分分析的降维方法的理解，尤其是具体的推导过程，只有通过实际的动手操作，才会发现很多的细节，比如说数据的中心化，以及相关参数计算的时候。这些都会让我们对方法更加熟练的掌握。

六、 参考文献

[1] 张学工. 模式识别[M]. 清华大学出版社, 2010.