第二次作业

李子龙
上海交通大学
计算机科学与工程系
logcreative@outlook.com

1 PCA

1.1 特征值分解

Algorithm 1: 特征值分解 PCA

Input: 数据集 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}, \mathbf{x}_t \in \mathbb{R}^{n\times 1}$
Output: 主成分 $\mathbf{w}$

计算平均值 $\mu \leftarrow \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i$;
foreach $i \leftarrow 1$ *to* N **do**
    $\mathbf{x}_i \leftarrow \mathbf{x}_i - \mu$;
计算散度矩阵 $C \leftarrow XX^T$;
特征值分解求 C 的特征值 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ 与对应的特征向量 $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$;
选取最大的特征值对应的特征向量与数据的乘积即为主成分 $\mathbf{w} \leftarrow \mathbf{v_1}^T\mathbf{X}$;
return $\mathbf{w}$;

优点

1. 简单易实现。
2. 解除线性相关。

缺点

1. 需要的内存大，需要先计算散度矩阵，当样本数量很大时，这一步消耗的时间复杂度比较高。
2. 计算散度矩阵这一步在数据量较少时可能会丢失精度。
3. 只能压缩一个方向（行或列）。

1.2 奇异值分解

Algorithm 2: 奇异值分解

Input: 数据集 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}, \mathbf{x}_t \in \mathbb{R}^{n\times 1}$

Output: 主成分 $\mathbf{w}$

计算平均值 $\mu \leftarrow \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i$;

foreach $i \leftarrow 1$ *to* N **do**

  $\mathbf{x}_i \leftarrow \mathbf{x}_i - \mu$;

奇异值分解 $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$;

两边同乘 $\mathbf{U}^T$，$\mathbf{U}^T\mathbf{X} = \mathbf{\Sigma}\mathbf{V}^T$ 得到压缩数据;

选取 $\mathbf{\Sigma}\mathbf{V}^T$ 中最大的那一个奇异值（习惯上应为左上角的值）对应的向量（一般为第一行）即为主成分 $\mathbf{w}$;

return $\mathbf{w}$;

优点

1. 可以直接对非方阵 $\mathbf{X}$ 进行奇异值分解，而特征值分解需要分解方阵 $\mathbf{X}\mathbf{X}^T$。可免去计算 $\mathbf{X}\mathbf{X}^T$ 的中间步骤。

2. 计算奇异值已经有快速地数值算法，在需要在时间空间与精度直接抉择时，可以选择后者直接取出较大的奇异值，精度上的折中是可以接受的。

3. 既能压缩行又能压缩列。

缺点

1. SVD 算法需要实现，算法实现难度比特征值分解大。

2. 分解后的矩阵缺少可解释性。

2 FA

$$
\begin{aligned}
p(\mathbf{y}|\mathbf{x}) &= \frac{p(\mathbf{x}|\mathbf{y})p(\mathbf{y})}{p(\mathbf{x})} && (2.1)\\
&= \frac{G(\mathbf{x}|\mathbf{A}\mathbf{y}+\mu, \mathbf{\Sigma}_e)G(\mathbf{y}|0, \mathbf{\Sigma}_y)}{p(\mathbf{A}\mathbf{y}+\mu+\mathbf{e})} && (2.2)\\
&= \frac{G(\mathbf{x}|\mathbf{A}\mathbf{y}+\mu, \mathbf{\Sigma}_e)G(\mathbf{y}|0, \mathbf{\Sigma}_y)}{G(\mathbf{y}|\mu+\mu_e, \mathbf{A}\mathbf{\Sigma}_y\mathbf{A}^T+\mathbf{\Sigma}_e)} && (2.3)
\end{aligned}
$$

公式 (2.1) 采用了贝叶斯规则，公式 (2.2) 采用了已知条件，公式 (2.3) 由下面的方式推导：

$$
\begin{aligned}
E(\mathbf{x}) &= E(\mathbf{Ay} + \mu + \mathbf{e}) \\
&= 0 + \mu + E(\mathbf{e}) \\
&= \mu + \mu_e \\
Cov(\mathbf{x}) &= Cov(\mathbf{Ay} + \mu + \mathbf{e}) \\
&= E((\mathbf{Ay} + \mu + \mathbf{e} - E(\mathbf{x}))(\mathbf{Ay} + \mu + \mathbf{e} - E(\mathbf{x}))^T) \\
&= E((\mathbf{Ay} + (\mathbf{e} - \mu_e))(\mathbf{Ay} + (\mathbf{e} - \mu_e))^T) \\
&= E(\mathbf{Ayy}^T\mathbf{A}^T + \mathbf{Ay}(\mathbf{e} - \mu_e) + (\mathbf{e} - \mu_e)(\mathbf{Ay})^T + (\mathbf{e} - \mu_e)(\mathbf{e} - \mu_e)^T) \\
&= \mathbf{A\Sigma}_y\mathbf{A} + \mathbf{\Sigma}_e
\end{aligned}
$$

而公式 (2.3) 就是答案。

3 ICA

4 FA 降维