模式识别作业 1-6

周圣威 1120213587

June 2024

1 作业1

已知权重矩阵 W（大小为 $n\times n$）和向量 Q（大小为 $n\times 1$），以及辅助矩阵 D。其中，D 是 W 的度矩阵（即 D 的对角元素是 W 每一行的元素和）。

首先，考虑表达式 $\sum_{i,j} w_{ij}(q_i-q_j)^2$。

展开该公式，得

$$\sum_{i,j} w_{ij}(q_i-q_j)^2=\sum_{i,j} w_{ij}(q_i^2-2q_iq_j+q_j^2)$$

将求和分解开，上式可变为

$$\sum_{i,j} w_{ij}q_i^2-2\sum_{i,j} w_{ij}q_iq_j+\sum_{i,j} w_{ij}q_j^2$$

由于上课老师并未提到 W 是否为对称矩阵，故做如下分类。

1.1 W 不是对称矩阵

我们可以合并这些项

$$\sum_i q_i^2\sum_j w_{ij}-2\sum_{i,j} w_{ij}q_iq_j+\sum_j q_j^2\sum_i w_{ij}$$

注意到 $\sum_j w_{ij}$ 实际上是 W 对应的度矩阵 D_1 的对角元素 d_i，$\sum_i w_{ij}$ 实际上是 W^T 对应的度矩阵 D_2 的对角元素 d_j，可得

$$\sum_i d_iq_i^2-2\sum_{i,j} w_{ij}q_iq_j+\sum_j d_jq_j^2$$

用矩阵的形式表示，可以表示为

$$Q^TD_1Q-2Q^TWQ+Q^TD_2Q$$

最终结果为

$$\sum_{i,j} w_{ij}(q_i-q_j)^2=Q^TD_1Q-2Q^TWQ+Q^TD_2Q$$

1.2 W 是对称矩阵

同样合并这些项，得

$$\sum_i q_i^2 \sum_j w_{ij} - 2\sum_{i,j} w_{ij} q_i q_j + \sum_j q_j^2 \sum_i w_{ij}$$

由于 W 是对称的矩阵（即 $w_{ij} = w_{ji}$），注意到第一项和第三项实际上是相同的，因为它们都表示 q_i 和 q_j 被权重和相应的项所权重后的平方和

$$\sum_i q_i^2 \sum_j w_{ij} - 2\sum_{i,j} w_{ij} q_i q_j + \sum_i q_i^2 \sum_j w_{ij}$$

进一步简化，注意到 $\sum_j w_{ij}$ 实际上是度矩阵 D 的对角元素 d_i

$$2\sum_i d_i q_i^2 - 2\sum_{i,j} w_{ij} q_i q_j$$

用矩阵的形式表示，可以表示为

$$2Q^T DQ - 2Q^T WQ$$

最终结果为

$$\sum_{i,j} w_{ij}(q_i - q_j)^2 = 2Q^T DQ - 2Q^T WQ$$

上课时老师说用 W、Q、D 表示，W 能推出 D，所以感觉应该是默认 W 是对称矩阵了。

2 作业2

考虑一个二分类问题，类的先验分布和似然概率分别如下：
先验分布： $P(\omega_1) = 0.6$、$P(\omega_2) = 0.4$
似然概率： 对于类 ω_1：

$$p(x|\omega_1) = \begin{cases} x & 0 \le x < 1 \\ 2 - x & 1 \le x \le 2 \\ 0 & \text{else} \end{cases}$$

对于类 ω_2：

$$p(x|\omega_2) = \begin{cases} x - 1 & 1 \le x < 2 \\ 3 - x & 2 \le x \le 3 \\ 0 & \text{else} \end{cases}$$

已知 MAP 分类准则的公式为

$$\text{MAP}(x) = \arg\max_i P(\omega_i|x)$$

根据贝叶斯公式，后验概率可以表示为：

$$P(\omega_i|x) = \frac{p(x|\omega_i)P(\omega_i)}{p(x)}$$

其中，$p(x)$ 是样本 x 的边际概率，可得

$$p(x) = p(x|\omega_1)P(\omega_1) + p(x|\omega_2)P(\omega_2)$$

当 $x = 1.6$ 时，有
似然概率为

$$p(1.6|\omega_1) = 2 - 1.6 = 0.4$$
$$p(1.6|\omega_2) = 1.6 - 1 = 0.6$$

边际概率为

$$p(1.6) = p(1.6|\omega_1)P(\omega_1) + p(1.6|\omega_2)P(\omega_2)$$
$$p(1.6) = 0.4 \times 0.6 + 0.6 \times 0.4 = 0.24 + 0.24 = 0.48$$

后验概率为

$$P(\omega_1|1.6) = \frac{p(1.6|\omega_1)P(\omega_1)}{p(1.6)} = \frac{0.4 \times 0.6}{0.48} = \frac{0.24}{0.48} = 0.5$$

$$P(\omega_2|1.6) = \frac{p(1.6|\omega_2)P(\omega_2)}{p(1.6)} = \frac{0.6 \times 0.4}{0.48} = \frac{0.24}{0.48} = 0.5$$

在这种情况下，后验概率是相同的，即 $P(\omega_1|1.6) = P(\omega_2|1.6)$，即我们无法根据MAP分类准则直接确定样本 $x = 1.6$ 的类别。因此可以选择一个策略，例如选择具有较高先验概率的类，即 ω_1。

所以，给定输入样本 $x = 1.6$，我们可以将其分类为类 ω_1。

3 作业3

已知随机变量 x 服从拉普拉斯分布，其概率密度函数为

$$f(x|\mu,\lambda) = \frac{1}{2\lambda}\exp\left(-\frac{|x-\mu|}{\lambda}\right)$$

同时已知尺度参数 λ。

首先找到使得似然函数最大的 μ 值。令似然函数 $L(\mu)$ 为

$$L(\mu) = \prod_{i=1}^{N} f(x_i|\mu,\lambda)$$

将拉普拉斯分布的概率密度函数代入：

$$L(\mu) = \prod_{i=1}^{N} \frac{1}{2\lambda}\exp\left(-\frac{|x_i-\mu|}{\lambda}\right)$$

取对数似然函数 $\ell(\mu)$：

$$\ell(\mu) = \ln L(\mu) = \sum_{i=1}^{N} \ln\left(\frac{1}{2\lambda}\exp\left(-\frac{|x_i-\mu|}{\lambda}\right)\right)$$

$$\ell(\mu) = \sum_{i=1}^{N}\left(\ln\left(\frac{1}{2\lambda}\right) - \frac{|x_i-\mu|}{\lambda}\right)$$

$$\ell(\mu) = N\ln\left(\frac{1}{2\lambda}\right) - \frac{1}{\lambda}\sum_{i=1}^{N}|x_i-\mu|$$

为了最大化 $\ell(\mu)$，我们需要最小化 $\sum_{i=1}^{N}|x_i-\mu|$，即找到这组数的中位数。即位置参数 μ 的最大似然估计是样本 x_i 的中位数。

所以，当已知尺度参数 λ 时，拉普拉斯分布位置参数 μ 的最大似然估计为样本数据的中位数。

4 作业4

首先初始化高斯成分的参数μ_c、σ_c和w_c，其中μ_c是均值，σ_c是协方差矩阵，w_c是每个成分c的混合系数。

在 E 步，计算每个高斯成分c对每个数据点x_i的权重

$$\gamma_{ic} = \frac{w(c)\mathcal{N}(x_i|\mu_c, \sigma_c)}{\sum_{c=1}^{C} w(c)\mathcal{N}(x_i|\mu_c, \sigma_c)}$$

其中$\mathcal{N}(x_i|\mu_c, \sigma_c)$是第$c$个高斯成分的概率密度函数。

在 M 步，使用在 E 步中计算的权重来更新参数，得

$$w(c) = \frac{N_c}{N}$$

$$\mu_c = \frac{1}{N_c} \sum_{i=1}^{N} \gamma_{ic} x_i$$

$$\sigma_c = \frac{1}{N_c} \sum_{i=1}^{N} \gamma_{ic} (x_i - \mu_c)(x_i - \mu_c)^T$$

其中$N_c = \sum_{i=1}^{N} \gamma_{ic}$是分配给簇$c$的有效点数。

最后我们进行评估，给定当前参数的数据的对数似然如下

$$\log L(\theta) = \sum_{i=1}^{N} \log \left(\sum_{c=1}^{C} w(c)\mathcal{N}(x_i|\mu_c, \sigma_c) \right)$$

如果对数似然的变化低于某个阈值，则算法已收敛。否则，返回到 E 步。

5 作业5

Algorithm 1 椭圆K-means算法

Require: 数据集 X, 簇的数量 K, 最大迭代次数 max_iters

Ensure: 簇中心 μ, 簇分配标签 $labels$

初始化:

  随机选择 K 个数据点作为初始簇中心 $\mu[k]$, $k = 1, \ldots, K$

  初始化每个簇的协方差矩阵 $S[k]$ 为单位矩阵

repeat

  簇分配步骤:

  for 每个数据点 $x[i]$ **do**

    计算 $x[i]$ 到每个簇中心 $\mu[k]$ 的椭圆距离 $d(x[i], \mu[k]) = \sqrt{(x[i] - \mu[k])^T S[k]^{-1}(x[i] - \mu[k])}$

    将 $x[i]$ 分配到最近的簇 j: $labels[i] = \arg\min_k d(x[i], \mu[k])$

  end for

  簇中心更新步骤:

  for 每个簇 k **do**

    更新簇中心 $\mu[k] =$ 簇 k 中所有数据点的均值

  end for

  协方差矩阵更新步骤:

  for 每个簇 k **do**

    计算新的协方差矩阵 $S[k] =$ 簇 k 中所有数据点的协方差矩阵

  end for

until 簇分配不再改变或达到最大迭代次数

return 簇中心 μ 和簇分配标签 $labels$

6 作业6

Algorithm 2 用竞争学习算法实现RBF网络的训练

Require: 数据集 X, 输出标签 Y, 隐藏层节点数目 K, 最大迭代次数 max_iters, 学习率 η

Ensure: 隐藏层中心 C, 隐藏层宽度 σ, 输出层权重 W

```
初始化:
   随机选择 K 个数据点作为初始中心 C[k], k = 1, ..., K
for 迭代次数 t 从 1 到 max_iters do
   for 每个输入数据点 x[i] do
      计算 x[i] 到每个中心 C[k] 的欧氏距离 d(x[i], C[k])
      找到最近的中心 j: j = arg min_k d(x[i], C[k])
      更新最近中心 j: C[j] = C[j] + η · (x[i] − C[j])
   end for
end for
计算宽度:
for 每个中心 C[k] do
   计算簇 k 中所有点到中心 C[k] 的平均距离
   设置 σ[k] = 平均距离
end for
调整输出层权重:
计算隐藏层输出矩阵 Φ, 其中 Φ[i, k] = exp(−||x[i] − C[k]||^2 / (2σ[k]^2))
使用最小二乘法求解权重 W: W = (Φ^T Φ)^{−1} Φ^T Y
return  隐藏层中心 C, 宽度 σ 和输出层权重 W
```

$$\Phi[i,k] = \exp\left(-\frac{||x[i]-C[k]||^2}{2\sigma[k]^2}\right), \quad W = (\Phi^T\Phi)^{-1}\Phi^T Y$$