

2017 年度《模式识别》课程作业

1. 假设 $p(\mathbf{x}|y = i) = N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, $i \in \{1, 2\}$, 即在一个 2 分类问题中, 每个类条件分布都是高斯分布, 他们虽然期望不同, 但协方差矩阵是一样的。假设两类的先验概率均为 0.5, 且使用 0-1 代价。
 - a) 写出 discriminant function g_1 和 g_2
 - b) 在保证分类结果不变的前提下, discriminant function 可以尽量简化, 写出你能得到的最简形式
 - c) 在二分类问题中, 可以将两个 discriminant function 简化为一个, 写出只用一个 discriminant function 时的分类规则。
2. 在 Matlab 中, 生成 lognormal 分布的 1000 个抽样 (其中 μ 和 σ 分布为 2, 0.5)
 - a) 在 matlab 文档中查找到合适的函数, 写出你生成抽样的 Matlab 语句。
 - b) 使用 ksdensity 函数的缺省参数进行无参数化密度估计, 画出其曲线。设置 width 参数分布为 0.2 和 4, 再得到两个图。简述你的体会。
 - c) 缺省参数中 width 由程序自动选择, 其值是多少? 若改为取 10000 个, 100000 个抽样, 其值又是多少? 趋势是什么? 解释为什么会出现这个趋势?
3. 假设本问题考虑的数据均服从高斯分布 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, 则马氏距离定义为 $d_M^2(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \mathbf{y})$ 。
 - a) 对数据进行白化变换, 将数据 \mathbf{x} 变成 \mathbf{x}' , 写出从 \mathbf{x} 变换到 \mathbf{x}' 的公式。
 - b) 证明 $d_M^2(\mathbf{x}, \mathbf{y}) = (\mathbf{x}' - \mathbf{y}')^T (\mathbf{x}' - \mathbf{y}')$
4.
 - a) KL divergence 不是一个 distance metric; 分别说明 metric 的四个条件中有哪几个满足, 哪几个不满足?
 - b) 证明不满足某个条件, 只需举出反例即可。考虑二值离散随机变量, 有三个分布分别为 $A=(1/2, 1/2)$, $B=(1/4, 3/4)$, $C=(1/8, 7/8)$ 。用 A、B、C 这三个分布举出反例来证明 a) 部分那些不满足的条件确实不成立。注意 b) 部分仅仅使用手工计算, 不允许使用计算器或程序。(hint: $\log_2 a + \log_2 b = \log_2 ab$, $\log_2 2^x = x$)
 - c) 用 Matlab 编程来计算上述反例中公式的数值, 验证你的手工计算的结果是对的。你的程序越简洁越好, 在你的答案中写出程序。
5.
 - a) 证明对任意的 $0 < x < \infty$, 有 $\ln x \leq x - 1$ 。等号何时成立?
 - b) 使用上述结论, 证明 $D_{KL}(p||q) \geq 0$ 总成立 (可以只针对离散随机变量进行证明)。
 - c) 等号何时成立? 说明理由。