2017-2018-1 CS28010 类脑智能期中考试

隋国新

2017 年 11 月 26 日

1 （视觉理论）请用自己的语言解释

(a) Wiesel-Hubel 特征检测理论

(b) Marr 的视觉计算理论

There are

2 （回归分析）感知机 perceptron

(a) 用自己的语言结合公式解释什么是 perceptron

(b) 请解释为什么用 perceptron 没有办法解决 XOR（异或）的求解

(c) 如何变化 perceptron 使之解决问题 b)?

(a) MP 神经元模型中，神经元接收到来自 n 个其他神经元传递过来的输入信号，这些输入信号通过带权重的连接进行传递，神经元接收到的总输入值将与神经元的阈值比较，通过激活函数处理产生神经元的输出。
perceptron 由两层神经元组成，输入层接收外界输入信号后传递给输出层，输出层是一个 MP 神经元。

(b) XOR 是一个非线性可分问题。perceptron 只有一层功能神经元，而 MP 神经元本质上是对输入进行线性处理后通过一个单调的激活函数，故不能解决这样的问题。

(c) 可以使用多层功能神经元。

3 （最小平方拟合）数据集 $x_i, i = 1, ..., n$，其中 x_i 是一个 d 维向量。假设它们的均值为 0。现有一个单位长向量 w。x_i 沿着向量 a_i 在 w 上的投影为 $\hat{x}_i$

(a) 请写出 $\hat{x}_i$ 关于 x_i, a_i 和 w 的表达式

(b) 求出使 $\Sigma_{i=1}^{n}||\hat{x} - x_i||$ 最小的 a_i 和 w

4 （高斯分布）$G(x|\mu_x, \sigma^2 I) = \frac{1}{(2\pi)^{d/2}\sigma d} exp(-\frac{1}{2\sigma^2}(x-\mu_x)^T(x-\mu_x)$，其中 x 为 d 维向量。$G(y|\mu_y, \Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} exp(-\frac{1}{2}(x-\mu_y)^T\Sigma^{-1}(x-\mu_y)$, 其中 y 为 d 维向量。X 和 y 之间满足 $x = By$

(a) 请写出 B 的表达式

(b) 请针对 B 表达式的形式来解释 B 的含义 (对 y 做了什么操作得到 x)

(a) 根据题意我们有：

$$
\begin{aligned}
E(x) &= E(By) \\
&= BE(y) \\
Cov(x) &= Cov(By) \\
&= E[(By - E(By))(By - E(By))^T] \\
&= E(Byy^TB^T - BE(y)y^TB^T - ByE(y)^TB^T + BE(y)E(y)^TB^T) \\
&= E(Byy^TB^T - B\mu_y\mu_y^TB^T) \\
&= B(E(yy^T) - \mu_y\mu_y^T)B^T \\
&= B\Sigma B^T
\end{aligned}
$$

即：$\begin{cases} \mu_x = B\mu_y \\ \sigma^2 I = B\Sigma B^T \end{cases}$

(b) 由 $\mu_x = B\mu_y$ 可知将 y 的均值移动至 x 的均值处，由 $\sigma^2 I = B\Sigma B^T$ 可知将 y 的方差压缩，对角线方向调整至 σ，反对角线方向分布密度与对角线方向一致。

5 （统计决策）请用自己的语言结合公式解释

(a) Bayesian classification

(b) Fisher discriminant analysis

(c) 它们的区别

(a) 根据贝叶斯定理，$P(c|x) = \frac{P(c)P(x|c)}{P(x)}$，其中 $P(c)$ 是分类为 c 的先验概率，$P(x|c)$ 是给定类标记后样本 x 的条件概率。Bayesian classification 就是根据这个公式，基于训练数据估计 $P(c)$ 和 $P(x|c)$，进而求出后验概率 $P(c|x)$。

(b) Fisher discriminant analysis

(c) 它们的区别

6 (高斯因子分析) 请用自己的语言结合公式解释 linear Gaussian factor analysis($X = Ay + e$) 的

(a) scale 不确定性

(b) rotation 不确定性

(c) additive 不确定性

(d) dimension 不确定性

(e) 若 factor y 的分布从高斯分布变成二项分布, 上述不确定性还存在么? 为什么?

定义 A 是一个 $n \times m$ 的矩阵, n 是 x 的维数, m 是 y 的维数，$m < n$

(a) scale 不确定性: y 服从于一个 m 维的高斯分布，均值与方差的 scale 都不确定，并且可以随着 A 的 scale 做出相应调整，存在 scale 不确定性

(b) rotation 不确定性：若不对 y 做特殊限制，y 可以旋转，存在 rotation 不确定性

(c) additive 不确定性: 考虑 Ay 与 e 的值的分配，二者的尺度都不确定时，additive 不确定性存在

(d) dimension 不确定性：y 的维度 m 不确定，存在 dimension 不确定性

(e) 若 factor y 的分布从高斯分布变成二项分布,scale 不确定性，rotation 不确定性，dimension 不确定性消失, 只有 additive 不确定性。

7 (最小平方聚类)

(a) 用自己的语言结合公式解释什么是最小平方聚类

(b) 最小平方聚类有三个无法解决的问题, 它们分别是什么? 为什么会存在?

(c) 就 k--means 算法而言, 在计算条件允许的情况下, 给定类别数 k, 如何求最优聚类?

(d) 但是因为代价太高, 一般会使用近似算法来求局部最优解。请用 EM 的思路来解释这个近似求解算法。(需要公式推导过程)

(a) 给定样本集 $x_i, i = 1, ..., m$, 聚类后获得簇划分 $C_i, i = 1, ..., k$，最小平方聚类方法最小化平方误差 $E = \sum_{i=1}^{k} \sum_{x \in C_i} ||x - \mu_i||_2^2$，E 值越小簇[illegible]样本相似度越高。

(b)
- 必须事先给出 K，而且对初值敏感，对于不同的初始值，结果可能不同
- 只能发现球状 Cluster，不适合于发现非凸形状的簇或者大小差别很大的簇
- 对噪声和孤立点数据敏感，如簇中含有异常点，将导致均值偏离严重。

(c) 每一个数据点的分类有 k 种，所有点的分类有 k^m 种，计算所有情况的平方误差，取最优解

(d) 假定样本以各自的簇为中心满足高斯分布，则可以使用 EM 混合高斯模型。根据 EM 算法，

- 选择初始的 K 个类别中心 $\mu_i, i = 1, ..., k$
- M-step: $\theta := argmax \sum_i \sum_{c^i} log \frac{p(x_i, c_i; \theta)}{Q_i(c^i)}$
  将软分类退化为硬分类，则有 $c_i := argmax(p(x_i, c_i; \theta))$
  不考虑高斯分布方差，则有 $c_i := argmin||x_i - \mu_i||^2$
  即对于每个样本 x_i，将其标记为距离类别中心最近的类别

- E-step: $Q_i(c^i) = p(c_i|x_i, \theta)$
  将软分类退化为硬分类后，不考虑概率，更新簇中心为 $\mu_j := \frac{\sum_{i=1}^{m} c_i = j x_i}{\sum_{i=1}^{m} c_i = j}$
- 重复前两步，直到类别中心的变化小于某阈值或者达到最大迭代次数