

复习大纲 2018 年 10 月 17 日

1. 什么是线性分类模型？什么是非线性分类模型？
2. 线性分类模型的判别函数形式？针对两类别/多类别分类，通常如何利用线性判别函数对任意观测样本进行类别决策？
3. 区分：决策域、决策边界、判别函数
4. 以两类别分类问题为例，不同维数的特征空间，分类边界？
5. 对于线性分类模型，特征空间任意观测点 \mathbf{x} 到分类边界的距离、代数距离？该点处判别函数取值？

$$g(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + \omega_0$$
$$\text{距离: } \|\mathbf{x} - \mathbf{x}_p\|_2 = \frac{|g(\mathbf{x})|}{\|\mathbf{w}\|}$$
$$\text{代数距离: } \mathbf{x} - \mathbf{x}_p = g(\mathbf{x}) \frac{\mathbf{w}}{\|\mathbf{w}\|^2}$$

1. 线性 Fisher 判别分类模型.

- (1) 线性 Fisher 判别分类模型构建的基本思想？
- (2) 针对两类别/多类别的分类，线性 Fisher 判别分析模型的准则函数形式？代表的意义？准则函数中各参数的意义？
- (4) 准则函数的最优值解形式？
- (5) 利用给定的训练样本集，面向两类别/多类别分类问题，设计 Fisher 判别分类模型，并基于该模型，对任意观测样本进行类别决策。实现步骤？
- (6) 如何基于线性 Fisher 判别分析模型，进行监督式特征提取？
- (7) 其它：
 - 给定已知类别标记的训练样本集，利用该样本集估算原始空间/投影空间的各类中心、类内散布矩阵(协方差矩阵)、总类内散布矩阵、类间散布矩阵。
 - 针对两类别/多类别分类问题，如何基于训练样本集，结合 Fisher's Discriminate Ratio(FDR, Fisher 判别比)对样本特征的类鉴别能力进行评价？

$$\text{两类问题: } FDR^{(i)} = \frac{(\mu_1^{(i)} - \mu_2^{(i)})^2}{(\sigma_1^{(i)})^2 + (\sigma_2^{(i)})^2}, \quad i = 1, 2, \dots, d$$
$$\text{多类问题:}$$
$$FDR^{(i)} = \sum_{k=1}^{C-1} \sum_{j>k}^C \frac{(\mu_k^{(i)} - \mu_j^{(i)})^2}{(\sigma_k^{(i)})^2 + (\sigma_j^{(i)})^2}, \quad i = 1, 2, \dots, d$$
$$FDR^{(i)} = \min_{\substack{k, j \in \{1, 2, \dots, C\} \\ \text{并且 } k < j}} \frac{(\mu_k^{(i)} - \mu_j^{(i)})^2}{(\sigma_k^{(i)})^2 + (\sigma_j^{(i)})^2}, \quad i = 1, 2, \dots, d$$
$$FDR^{(i)} = \frac{\sum_{j=1}^C (\mu_j^{(i)} - \mu^{(i)})^2}{\sum_{j=1}^C (\sigma_j^{(i)})^2}, \quad i = 1, 2, \dots, d$$

2. 感知器分类模型.

- (1)什么是线性可分的训练样本集?
- (2)感知器分类模型的判别函数形式?
- (3)设计感知器分类模型的基本思想? 模型学习, 构建的准则函数形式? 准则函数代表的意义? 准则函数中各参数的意义?
- (4)准则函数最优值解求取方法(梯度下降法)

3. 线性/核 SVM(非线性 SVM)

- (1)以 C-SVM 分类模型为例, 线性/核 SVM 分类模型的设计思想.
- (2)设计线性 C-SVM 分类模型时, 原始目标函数及约束条件?目标函数的意义? 各参数意义?
- (3)原始问题如何转化为对偶问题?对偶问题的目标函数及约束条件?
- (4)线性 C-SVM 分类模型的判别函数形式? 判别函数的各参数与训练样本关系?
- (5)如何将训练样本集区分为: 支持向量(错分支持向量、边界支持向量)、非支持向量?
- (6)线性 C-SVM 分类边界方程? H_1 、 H_2 的方程? H_1 、 H_2 与支持向量的位置关系?
- (7)SVM 是大间隔分类模型。分类间隔=?

- (8)对于核 SVM 分类模型, 原始目标函数及约束条件?目标函数的意义? 各参数意义? 对偶目标函数及约束条件?
- (9) 核 C-SVM 分类模型的判别函数形式? 判别函数的各参数与训练样本关系? 分类边界的方程? H_1 、 H_2 的方程?

- (10) 以 C-SVM 分类模型为例, 给定已知类别标记的训练样本集, 描述基于该样本集, 进行核 SVM 分类模型的实现步骤; 若完成了分类模型的学习, 如何基于得到的分类模型, 进行特征空间任意观测样本的类别决策?
- (11) 若上述分类模型采用 RBF 核函数,以“grid searching”法为例,如何基于 K-fold 交叉验证法,实现 C-SVM 分类模型的超参数优选? 简述参数优选步骤.

- (12)多类别 SVM. 给定已知类别标记的训练样本集, 如何基于两类别的 SVM 分类模型, 构建多类别的分类模型(one vs. one, one vs. rest)? 给定特征空间任意观测样本, 如何对样本的类别进行预测?