HW6

PB20111689 蓝俊玮

6.4

> 试讨论线性判别分析与线性核支持向量机在何种条件下等价

线性判别分析能够解决 `n` 分类问题，而线性核 SVM 只能解决二分类问题。

从求解的角度来考虑，线性判别分析的目标是要让同类的投影点尽可能接近，而异类样例的投影点尽可能远离。而考虑 SVM，SVM 求解出来的超平面会使这两类样例尽可能远。当线性判别分析的投影向量和线性核 SVM 的超平面向量垂直的时候，可以看出，SVM 的最大间隔就是线性判别分析所要求的异类投影点间距，同时在这种情况下，线性判别分析的同类样例的投影点也会被这个超平面所划分在一起，使其间隔较小。

所以如果线性判别分析求解出来的投影向量 w_1 和 线性核 SVM 求解出来的超平面向量 w_2 垂直的时候，以及在数据集只有两类的时候和数据集时线性可分的情况下，两种是等价的。

6.6

> 试分析 SVM 对噪声敏感的原因

SVM 的基本形态是一个硬间隔分类器，它要求所有样本都满足硬间隔约束。因此噪声很容易影响 SVM 的学习。同时存在噪声时，SVM 容易受噪声信息的影响，将训练得到的超平面向两个类间靠拢，导致训练的泛化能力降低，尤其是当噪声成为支持向量时，会直接影响整个超平面。并且当 SVM 推广到到使用核函数时，会得到一个更复杂的模型，此时噪声也会一并被映射到更高维的特征，可能会对训练造成更意想不到的结果。因此 SVM 对噪声敏感。

6.9

> 试使用核技巧推广对率回归，产生“核对率回归”

原始对率回归问题为

$$\min_{\beta} \quad E = l(\boldsymbol{\beta}) = \sum_{i=1}^{m}(-y_i\boldsymbol{\beta}^T\hat{\boldsymbol{x}}_{\boldsymbol{i}} + \ln\left(1+e^{\boldsymbol{\beta}^T\hat{\boldsymbol{x}}_{\boldsymbol{i}}}\right)) \quad (3.27)$$

则由表示定理，可以写出

$$h(\boldsymbol{x}) = \sum_{i=1}^{m}\alpha_i\boldsymbol{\kappa}(x, x_i)$$

再有

$$h(\boldsymbol{x}) = \boldsymbol{\beta}^T\phi(\boldsymbol{x})$$

得到

$$\boldsymbol{\beta} = \sum_{i=1}^{m}\alpha_i\phi(\boldsymbol{x})$$

则可以得到

$$\min_{\alpha} \quad E = \sum_{i=1}^{m}(-y_i\sum_{i=1}^{m}\alpha_i\phi(\boldsymbol{x})^T\phi(\boldsymbol{x}_{\boldsymbol{i}}) + \ln\left(1+e^{\sum_{i=1}^{m}\alpha_i\phi(\boldsymbol{x})^T\phi(\boldsymbol{x}_{\boldsymbol{i}})}\right)$$

$$\min_{\alpha} \quad E = \sum_{i=1}^{m}(-y_i\sum_{i=j}^{m}\alpha_j\boldsymbol{\kappa}(x, x_j) + \ln\left(1+e^{\sum_{j=1}^{m}\alpha_j\boldsymbol{\kappa}(x,x_j)}\right))$$

$$\min_{\alpha} \quad E = \sum_{i=1}^{m}(-y_i\sum_{i=j}^{m}\alpha_jK_j + \ln\left(1+e^{\sum_{j=1}^{m}\alpha_jK_j}\right))$$

T4

设原式中的 $\boldsymbol{\alpha} = \tilde{\boldsymbol{\alpha}}$

则设

$$\boldsymbol{\alpha} = \begin{pmatrix} \tilde{\boldsymbol{\alpha}} \\ \hat{\boldsymbol{\alpha}} \end{pmatrix} \quad \boldsymbol{v} = \begin{pmatrix} -y-\varepsilon \\ y-\varepsilon \end{pmatrix}$$

同时

$$\tilde{\boldsymbol{\alpha}}^T K \tilde{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}}^T K \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}}^T K \tilde{\boldsymbol{\alpha}} + \hat{\boldsymbol{\alpha}}^T K \hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}^T \begin{pmatrix} \boldsymbol{\kappa} & -\boldsymbol{\kappa} \\ -\boldsymbol{\kappa} & \boldsymbol{\kappa} \end{pmatrix} \boldsymbol{\alpha}$$

则设

$$\boldsymbol{K} = \begin{pmatrix} \boldsymbol{\kappa} & -\boldsymbol{\kappa} \\ -\boldsymbol{\kappa} & \boldsymbol{\kappa} \end{pmatrix}$$

最后设

$$\boldsymbol{u} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

则根据这些 $\boldsymbol{\alpha}, \boldsymbol{v}, \boldsymbol{K}, \boldsymbol{u}$，可以将支持向量回归的对偶问题转化为如下标准形式：

$$\max_{\boldsymbol{\alpha}} g(\boldsymbol{\alpha}) = \boldsymbol{\alpha}^T \boldsymbol{v} - \frac{1}{2} \boldsymbol{\alpha}^T \boldsymbol{K} \boldsymbol{\alpha}$$
$$s.t. C \geq \boldsymbol{\alpha} \geq 0 \ and \ \boldsymbol{\alpha}^T \boldsymbol{u} = 0$$

T5

取平方项默认的映射空间大小：

$$\boldsymbol{\kappa}(x_i, x_j) = \phi(x_i)^T \phi(x_j) = (x_i^T x_j)^2 = (x_{i1}x_{j1} + x_{i2}x_{j2} + \ldots + x_{in}x_{jn})^2 = \sum_{k=1}^{k=n} (x_{ik}x_{jk})^2 + 2 \times \sum_{k=1}^{k=n} \sum_{l \neq k} x_{ik}x_{jk}x_{il}x_{jl}$$

$$\Rightarrow \quad \phi(x_i) = (\frac{\sqrt{2}}{\sqrt{i_1! i_2! \cdots i_n!}} x_{i1}^{i_1} x_{i2}^{i_2} \cdots x_{in}^{i_n}) \quad i_1 + i_2 + \cdots + i_n = 2$$

即有：

$$\phi(x_i) = (x_{i1}^2,\ \sqrt{2}x_{i1}x_{i2},\ \sqrt{2}x_{i1}x_{i_3},\ \ldots,\ x_{i2}^2,\ \ldots,\ x_{in}^2,\ \ldots)$$