

深入理解SVM

七月在线
讲师 老冯

课程回顾

□ 在线课程，听什么？

- 听 **Big Ideas**，细节课后复习
- 数学公式不是**big ideas**，而是**big ideas**的描述方法之一，好的课程会在数学公式之前先介绍**intuition**（动机）

□ 学数学有什么用？

- 知道什么问题是可以数学解决的
- （偶尔）知道如何去解决



机器学习的四个步骤

- 特征表示
- 建立模型
- 设立目标函数
- 优化求解



监督学习和决策边界



SVM的核心思想一：最大间隔



SVM的核心思想二：决策公式



SVM的核心思想三： 目标函数





What's next?

高斯？

拉普拉斯？

拉格朗日？



SVM的核心思想四：优化理论





代入决策公式



SVM的终极杀器：核方法



SMO





Hinge loss



Regularization

All the answers so far are of the form

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

They require the inversion of $\mathbf{X}^T \mathbf{X}$. This can lead to problems if the system of equations is poorly conditioned. A solution is to add a small element to the diagonal:

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^T \mathbf{X} + \delta^2 I_d)^{-1} \mathbf{X}^T \mathbf{y}$$

This is the ridge regression estimate. It is the solution to the following **regularised quadratic cost function**

$$J(\boldsymbol{\theta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) + \delta^2 \boldsymbol{\theta}^T \boldsymbol{\theta}$$

Derivation

$$\begin{aligned}\frac{\partial}{\partial \theta} J(\theta) &= \frac{\partial}{\partial \theta} \left\{ (y - x\theta)^T (y - x\theta) + s^2 \theta^T \overset{\substack{\text{identity} \\ \text{matrix}}}{I} \theta \right\} \\ &= \frac{\partial}{\partial \theta} \left\{ y^T y - 2y^T x \theta + \theta^T x^T x \theta + \theta^T (s^2 I) \theta \right\} \\ &= -2x^T y + 2x^T x \theta + 2s^2 I \theta \\ &= -2x^T y + 2(x^T x + s^2 I) \theta\end{aligned}$$

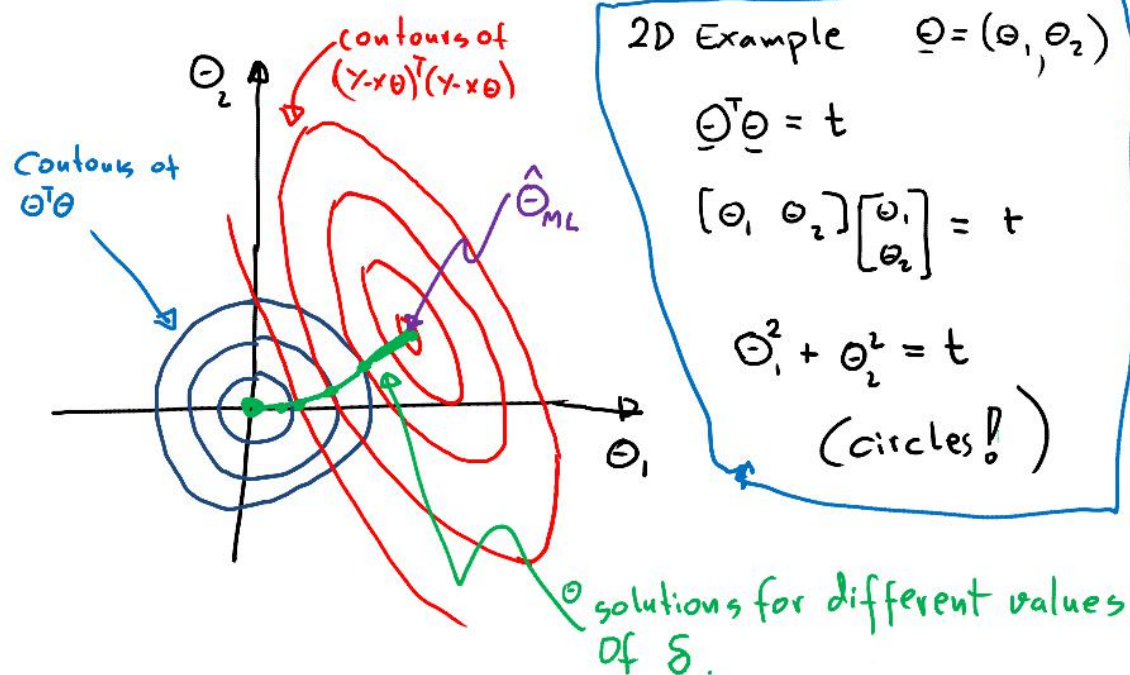
Equating to zero, yields

$$\hat{\theta}_{\text{ridge}} = (x^T x + s^2 I)^{-1} x^T y$$

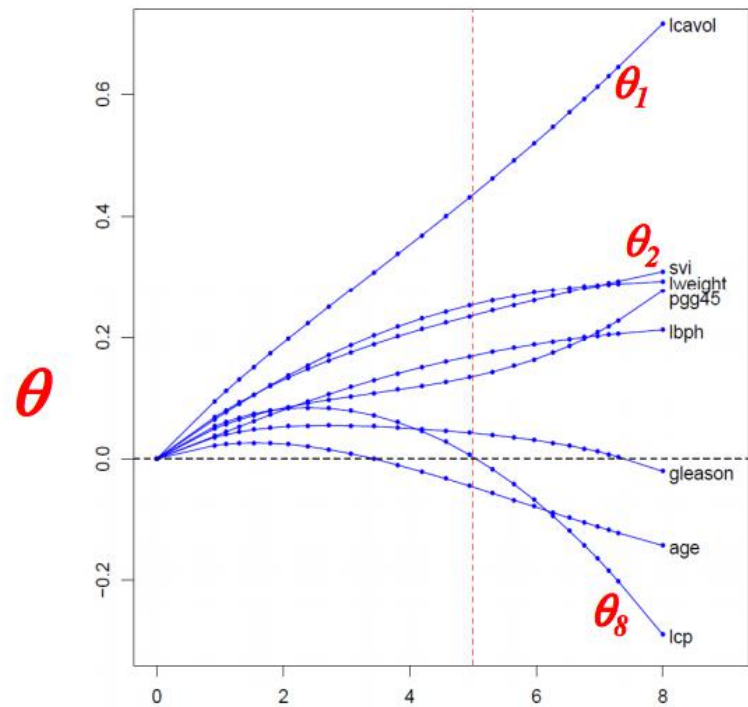
Ridge regression as constrained optimization

$$J(\theta) = (y - X\theta)^T(y - X\theta) + \delta^2\theta^T\theta$$

$$\theta : \min_{\theta^T\theta \leq t(\delta)} \{(y - X\theta)^T(y - X\theta)\}$$



As δ increases, $t(\delta)$ decreases and each θ_i goes to zero.



感谢大家么么哒！

记得报名其他班啊。