Spring 2016

Problem 2 Given training samples $\{(x^{(1)}, y^{(1)}), \ldots, (x^{(n)}, y^{(n)})\}$, ridge regression seeks to predict each response $y^{(i)}$ with a linear model $\theta \cdot x^{(i)}$ while encouraging θ to have a small norm. We omit the offset parameter for simplicity. Specifically, θ is estimated by minimizing

$$\left[\frac{1}{n}\sum_{i=1}^{n}(y^{(i)} - \theta \cdot x^{(i)})^2/2\right] + \frac{\lambda}{2}\|\theta\|^2, \tag{6}$$

where $\lambda \geq 0$ is the regularization parameter, typically chosen in advance.

(2.1) **(2 points)** What is the solution $\hat{\theta}$ that minimizes Eq(6) if $\lambda \to \infty$?

(2.2) **(2 points)** If we assume that $(1/n)\sum_{i=1}^{n}(y^{(i)})^2/2 = 1$, sketch in the figure how the squared training error

$$\frac{1}{n}\sum_{i=1}^{n}(y^{(i)} - \hat{\theta}(\lambda) \cdot x^{(i)})^2/2 \tag{7}$$

behaves as we vary λ. Here $\hat{\theta}(\lambda)$ is the solution to Eq(6) with the chosen λ.