Chapter *2* Linear Model

Siheng Zhang
zhangsiheng@cvte.com

December 12, 2020

This part corresponds to , and mainly answers the following questions:

-

Contents

1 Linear classification

2 Linear regression

$$\min_{\mathbf{w}} \sum_{i=1}^{N} (\mathbf{w}\mathbf{x}_i - y_i)^2 \tag{1}$$

2.1 MLE

Suppose that y_i is the , Gaussian noise, $y_i - \mathbf{w}\mathbf{x}_i \sim \mathcal{N}(0, \sigma^2)$, the log likelihood function is

$$\log \mathcal{L} = -\frac{N}{2} \log 2\pi - \sum_{i=1}^{N} \frac{(y_i - \mathbf{w}\mathbf{x}_i)^2}{2} \tag{2}$$

Obviously, MLE is equivalent to linear regression.

2.2 Ridge

Prior $\mathbf{w} \sim \mathcal{N}(0, \mathbf{I})$, posterior

$$p(\mathbf{w}|S) \propto p(\mathbf{w})p(S|\mathbf{w}) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}\mathbf{w}^\top \mathbf{w}\right) \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - \mathbf{w}\mathbf{x}_i)^2}{2}\right) \tag{3}$$

Maximizing the log posterior function is equivalent to the ridge regression.

2.2.1 Ridge in the MAP viewpoint

2.3 Lasso

2.3.1 Lasso in the MAP viewpoint