Deep Learning - Theory and Practice

Linear Regression, Least Squares
Classification and Logistic Regression

13-02-2020

http://leap.ee.iisc.ac.in/sriram/teaching/DL20/

deeplearning.cce2020@gmail.com

Established
1911

Linear Regression

$$y(\mathbf{x}, \mathbf{w}) = \sum_{j=0}^{M-1} w_j \phi_j(\mathbf{x}) = \mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x})$$

$$t = y(\mathbf{x}, \mathbf{w}) + \epsilon$$

- Solution to Maximum Likelihood problem is the least squares solution

$$\nabla \ln p(\mathbf{t}|\mathbf{w}, \beta) = \sum_{n=1}^{N} \left\{ t_n - \mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}_n) \right\} \boldsymbol{\phi}(\mathbf{x}_n)^{\mathrm{T}}.$$

Pseudo Inverse Based Solution

Bishop - PRML book (Chap 3)

Established 1911

INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

Choice of Basis Functions

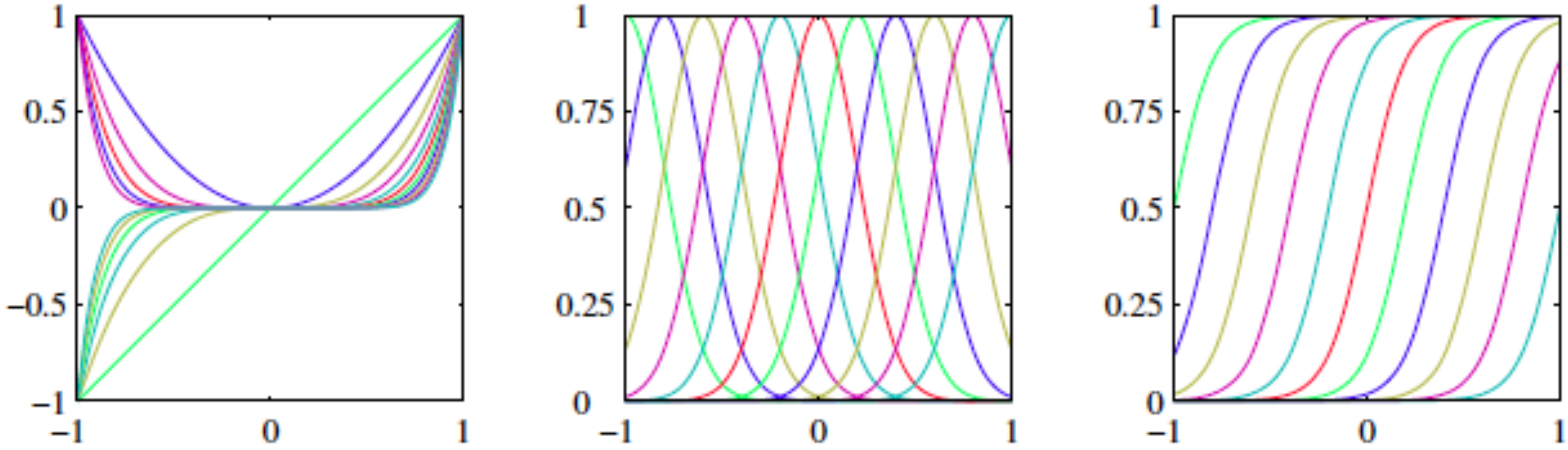
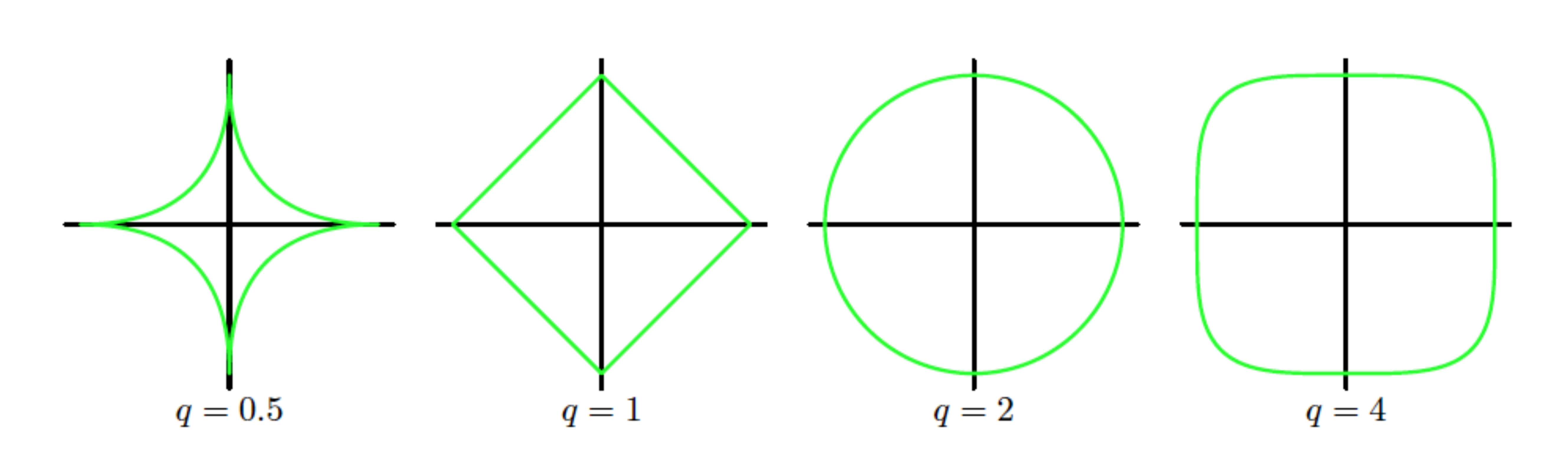


Figure 3.1 Examples of basis functions, showing polynomials on the left, Gaussians of the form (3.4) in the centre, and sigmoidal of the form (3.5) on the right.

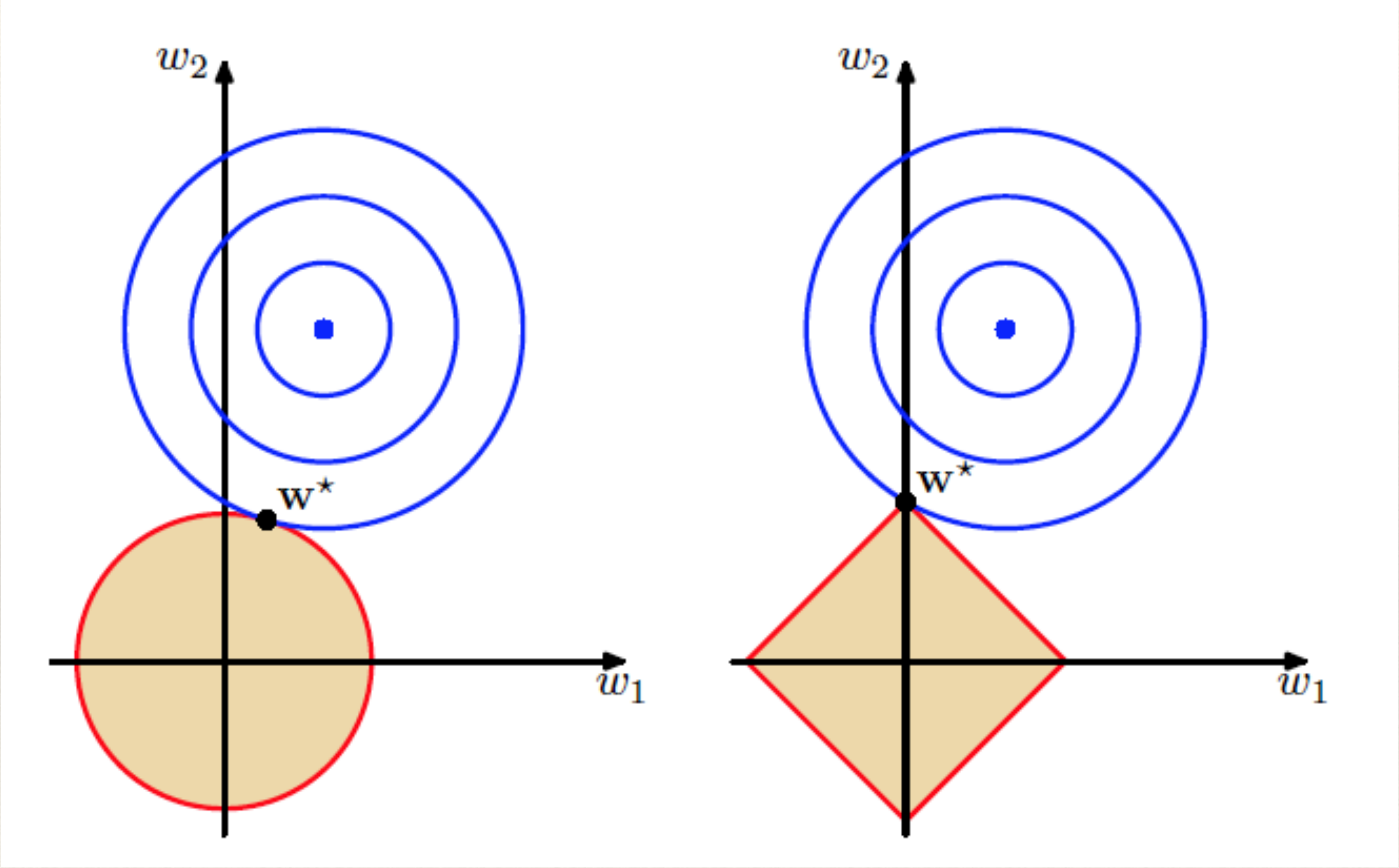
Regularized Least Squares

- Optimize a modified cost function $E_D(\mathbf{w}) + \lambda E_W(\mathbf{w})$



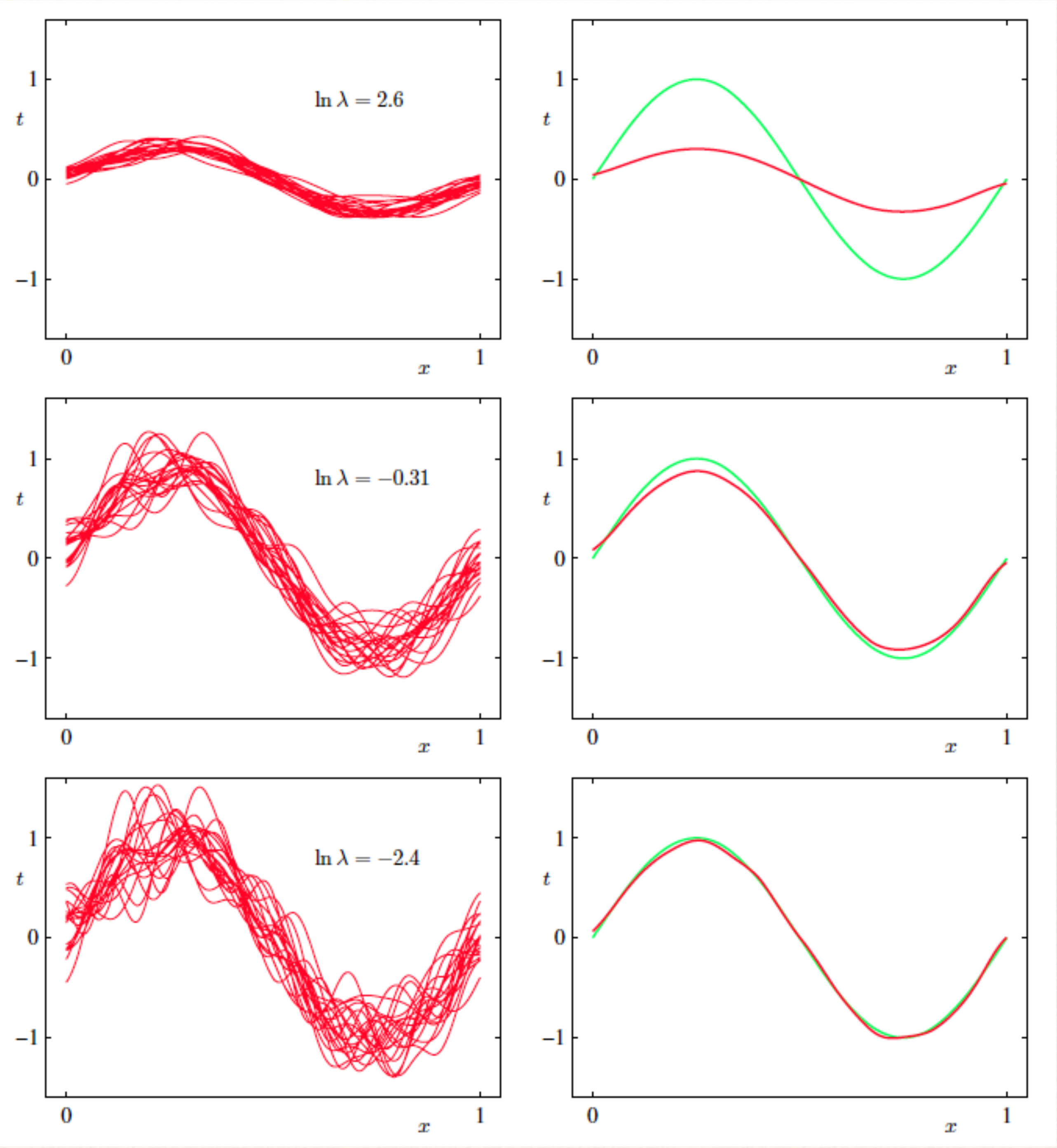
Bishop - PRML book (Chap 3)

Regularized Least Squares



Bishop - PRML book (Chap 3)

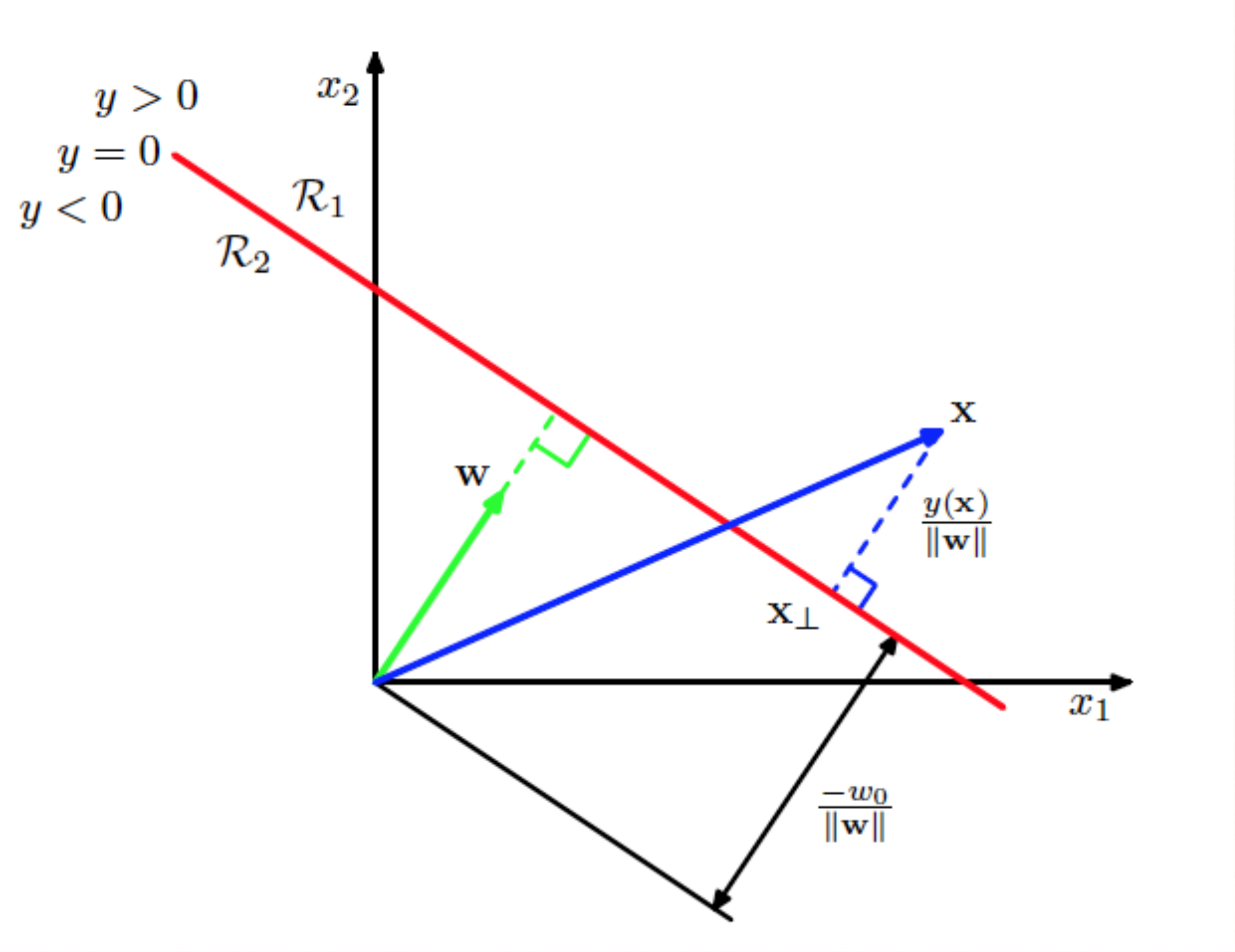
Choice of Regularization Parameter



Linear Models for Classification

- Optimize a modified cost function

$$y(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x} + w_0$$



Least Squares for Classification

- K-class classification problem

$$y_k(\mathbf{x}) = \mathbf{w}_k^{\mathrm{T}}\mathbf{x} + w_{k0}$$

$$\mathbf{y}(\mathbf{x}) = \widetilde{\mathbf{W}}^{\mathrm{T}}\widetilde{\mathbf{x}}$$

- With 1-of-K hot encoding, and least squares regression

$$E_D(\widetilde{\mathbf{W}}) = \frac{1}{2}\mathrm{Tr}\left\{(\widetilde{\mathbf{X}}\widetilde{\mathbf{W}} - \mathbf{T})^{\mathrm{T}}(\widetilde{\mathbf{X}}\widetilde{\mathbf{W}} - \mathbf{T})\right\}$$

Logistic Regression

- 2- class logistic regression $p(\mathcal{C}_1|\phi) = y(\phi) = \sigma\left(\mathbf{w}^{\mathrm{T}}\phi\right)$

- Maximum likelihood solution $\nabla E(\mathbf{w}) = \sum_{n=1}^{N}(y_n - t_n)\phi_n$

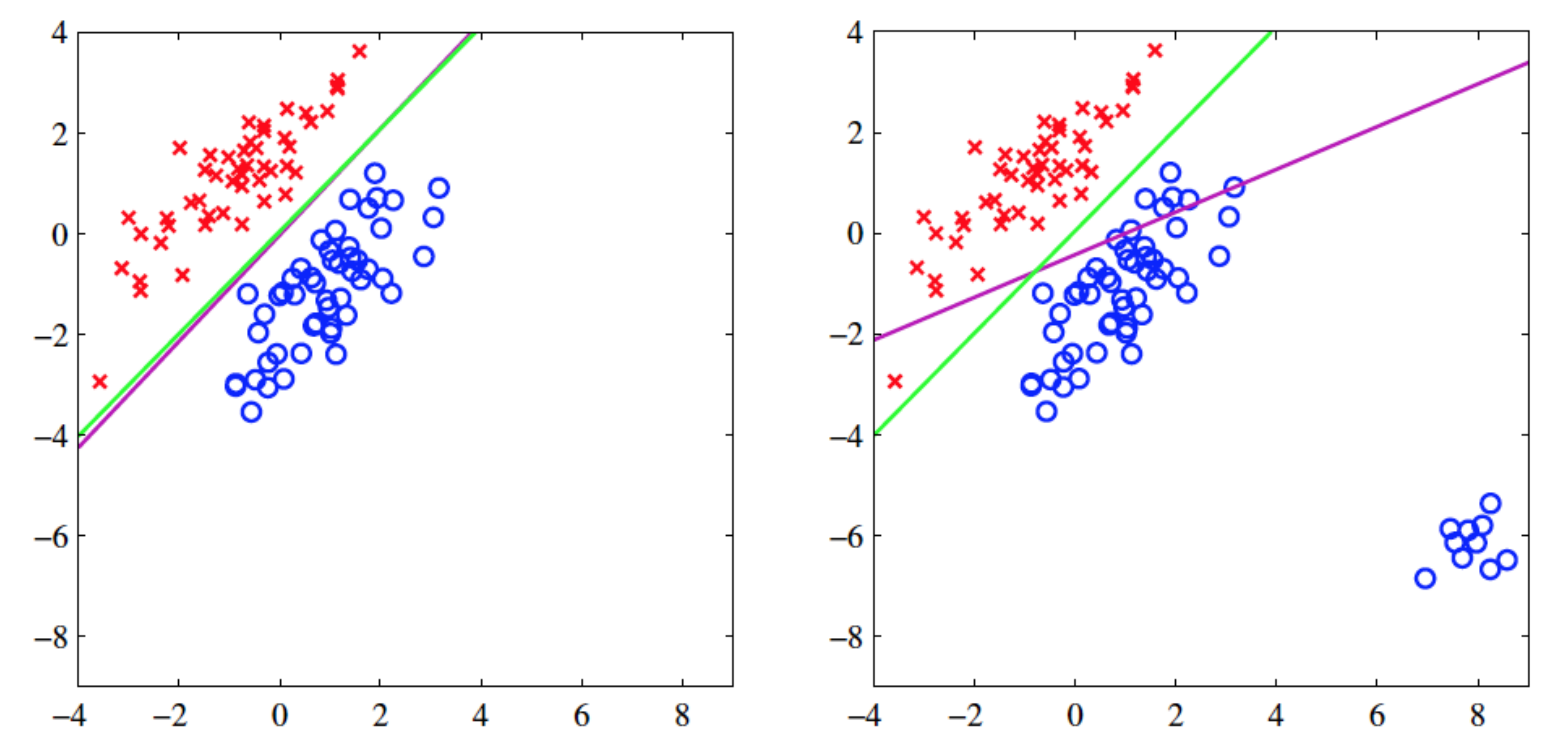
- K-class logistic regression $$p(\mathcal{C}_k|\phi) = y_k(\phi) = \frac{\exp(a_k)}{\sum_j \exp(a_j)}$$

$$a_k = \mathbf{w}_k^{\mathrm{T}}\phi.$$

- Maximum likelihood solution

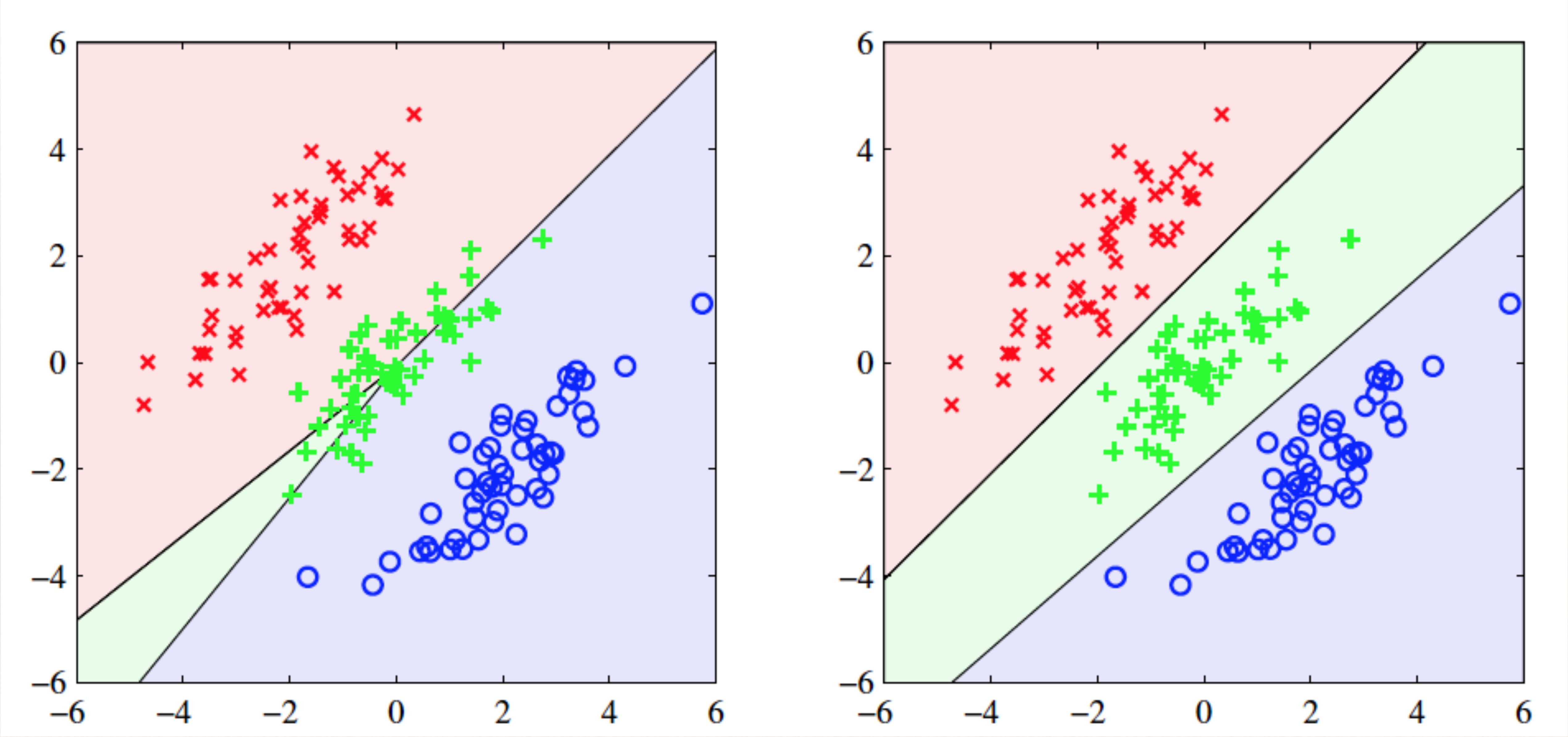
$$\nabla_{\mathbf{w}_j} E(\mathbf{w}_1, \ldots, \mathbf{w}_K) = \sum_{n=1}^{N}(y_{nj} - t_{nj})\,\phi_n$$

Least Squares versus Logistic Regression



Bishop - PRML book (Chap 4)

Least Squares versus Logistic Regression



Bishop - PRML book (Chap 4)