



**Figure 4.6** Plots of the quantity  $L_q = |f - t|^q$  for various values of  $q$ .

Section 6.5

of Gaussians to give multimodal conditional distributions, which often arise in the solution of *inverse problems*. Our focus in this section has been on decision theory for regression problems, and in the next chapter we shall develop analogous concepts for classification tasks.

Section 5.2

### 4.3. The Bias–Variance Trade-off

Section 1.2

So far in our discussion of linear models for regression, we have assumed that the form and number of basis functions are both given. We have also seen that the use of maximum likelihood can lead to severe over-fitting if complex models are trained using data sets of limited size. However, limiting the number of basis functions to avoid over-fitting has the side effect of limiting the flexibility of the model to capture interesting and important trends in the data. Although a regularization term can control over-fitting for models with many parameters, this raises the question of how to determine a suitable value for the regularization coefficient  $\lambda$ . Seeking the