6.2 Weighting data

(i) For Gaussian error distributions, the sum of squares, weighted by the population variances, turns up naturally in maximum likelihood estimation. Since maximum likelihood has the various good properties listed in the Chapter, we could reasonably describe this as the "optimum" weighting.
A large class of error distributions is of the form

$$f\left(\frac{x-\mu}{\beta}\right)$$

where μ is a location parameter and β is a scale parameter. Clearly a maximum-likelihood attack will yield a product of error terms weighted by the relevant $1/\beta$, in some way which depends on the form of f. For instance, for an exponential distribution, the weight is just $1/\beta$. It is easy to see (from the definition of variance) that the variance of the distribution f must be proportional to β^2. So in an exponential error distribution the weight goes inversely with the standard deviation and not with the variance.
(i) If the errors are not known to be Gaussian, another approach is possible. We can do something with linear estimators; that is, we assume that our parameter α can be estimated by a linear function of the data X_i, as follows:

$$\hat{\alpha} = \sum_i w_i X_i$$

where the w_i are the weights we want. Taking another meaning of "optimum" we would like this weighting to minimize the variance in our estimator,

$$\text{var}[\hat{\alpha}] = \sum_i w_i^2 \text{var}[X_i]$$

so that, using δ to denote a small variation

$$0 = \sum_i w_i \, \delta w_i \text{var}[X_i].$$

Also, we have the constraint that the weights should be normalized, $\sum_i w_i = 1$, so that

$$0 = \sum_i \delta w_i.$$

These two variational equations can be solved by the standard method of Lagrange multipliers, yielding the result

$$w_i = \left(\frac{1}{1/\sum_i \text{var}[X_i]}\right) \frac{1}{\text{var}[X_i]}.$$

Note that in approach (i), the weight is the population variance, whereas in approach (ii) the weight is the sample variance.