- Page 56, line 17 from bottom should read: "..... we will adhere to the squared error loss function,"

- Page 56, line 14 from bottom should read: "The squared error loss function ....."

- Page 56, line 10 from bottom should read: "then, the squared error loss function ....."

- Page 56, line 6 from bottom should read: "The combined choice of linearity with the squared error loss function turns out ..."

- Page 58, line 8 from top should read: "Using the LS cost function ....."

- Page 59, caption of Figure 3.3 should read: "The least-squares cost function ...."

- Page 64, the first equation reading from top should be:

$$P(y|\boldsymbol{x}) = \frac{p(y,\boldsymbol{x})}{p(\boldsymbol{x})} = \frac{p(y,\boldsymbol{x})}{\sum_y p(y,\boldsymbol{x})}.$$

- Page 66, line 17 from top should read: "Let us denote by $\hat{\theta}_{\mathrm{MVU}}$ a solution of the task $\min_{\theta:\ \mathbb{E}[\theta]=\theta_o} \mathrm{MSE}(\theta)$. ...."

- Page 67. In the last paragraph of Section 3.5.1, to comply with the previously used notation, θ should be replaced by $\hat{\theta}$. Thus the paragraph should read as: "Note that what we said so far is readily generalized to parameter vectors. An unbiased parameter vector estimator satisfies

$$\mathbb{E}[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}_o,$$

and the MSE around the true value, $\boldsymbol{\theta}_o$, is defined as

$$\mathrm{MSE} = \mathbb{E}\left[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o)^T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o)\right].$$

Looking carefully at the previous definition reveals that the MSE for a parameter vector estimator is the sum of the MSEs of the components, $\hat{\theta}_i$, $i = 1, 2 \ldots, l$, around the corresponding true values θ_{oi}.

- Page 72. To comply with the notation used in the rest of the chapters, the colon should be deleted in both Eqs. (3.37) and (3.38), that is

$$\text{minimize} \quad J(\boldsymbol{\theta}) = \sum_{n=1}^{N} \left(y_n - \boldsymbol{\theta}^T \boldsymbol{x}_n\right)^2, \tag{3.37}$$

$$\text{subject to} \quad \|\boldsymbol{\theta}\|^2 \leq \rho, \tag{3.38}$$

- Page 73, the same as before for Eq. (3.39), i.e.

$$\boxed{\text{minimize} \qquad L(\boldsymbol{\theta}, \lambda) = \sum_{n=1}^{N} \left(y_n - \boldsymbol{\theta}^T \boldsymbol{x}\right)^2 + \lambda \|\boldsymbol{\theta}\|^2 : \quad \text{Ridge Regression.}} \tag{3.39}$$

- Page 74, line 14 from bottom should read: "this data dependence of the task; that is, ..."

- Page 78, line 4 from bottom, it should read: "the mean-square deviation from the optimal estimate comprises two terms....."

- Page 84, the first paragraph of Section 3.11 refers to a single parameter, although the correct one should be to refer to many. Hence, the paragraph should read:
"In our discussion, so far, we have assumed that the parameters associated with the functional form of the adopted model are deterministic constants, whose values are unknown to us. In this section, we will follow a different rationale. The unknown parameters will be treated as random variables. Hence, whenever our goal is to estimate their values, this is conceived as an effort to estimate the values of a *specific* realization that corresponds to the observed data....."

- Page 101, line 15 from top should read: "conditioned on $\mathcal{X}$, turns out to be Gaussian with mean μ_N and variance σ_N^2, where ...."

- Page 101, line 16 from top should read:

$$\mu_N := \frac{N\sigma_0^2 \bar{x} + \sigma^2 \mu_0}{N\sigma_0^2 + \sigma_\eta^2}, \quad \sigma_N^2 := \frac{\sigma^2 \sigma_0^2}{N\sigma_0^2 + \sigma_\eta^2}.$$