Exercise Sheet #6

due on Monday, 6 December 2021, 10am sharp

1. **EXAMple: Linear Least-Squares** Consider a data-set in which each data point t_n has a weighting $r_n > 0$, so that the sum-of-square error function is

$$E(\omega) = \sum_{n=1}^{N} r_n (t_n - \omega^\top x_n)^2. \tag{1}$$

   (a) Find the parameter-vector $\hat{\omega}$ which minimizes this error function.

   (b) How could you use this error function to fit data in which some measurements were repeated several times, producing the exact same result? What would you use for each r_n in that case?

   (c) How could you use this error function to fit data in which each measurement had a different, known value for the observation noise σ^2? What would you use for each r_n in that case?

2. **Theory Question: Bayesian Linear Regression** We have data $D = \{(\boldsymbol{x}_1, t_1), \ldots, (\boldsymbol{x}_N, t_N)\}$, and want to model it with a linear regression model by $t \approx y(x, \omega) + \epsilon$, where ω is M-dimensional, and $y(x, \omega) = \omega^\top x$. We assume that noise ϵ is independent, identically distributed and Gaussian,

$$\epsilon \sim \mathcal{N}(0, \beta^{-1}) \tag{2}$$
$$t|\boldsymbol{x}, \omega, \beta \sim \mathcal{N}(y(\boldsymbol{x}, \omega), \beta^{-1}). \tag{3}$$

   We use a multivariate Gaussian as a prior on the parameters ω,

$$\omega_i \sim \mathcal{N}(0, \alpha^{-1}) \tag{4}$$
$$p(\omega|\alpha) = \prod_{i=1}^{M} \sqrt{\frac{\alpha}{2\pi}} \exp\left(-\frac{\alpha}{2}\omega_i^2\right) \tag{5}$$
$$= \left(\frac{\alpha}{2\pi}\right)^{M/2} \exp\left(-\frac{\alpha}{2}\omega^\top\omega\right) \tag{6}$$

   We assume that the noise variance β^{-1} and the prior variance α^{-1} are known. Verify that the posterior distribution over ω is $\mathcal{N}(\mu, \Sigma)$, where

$$\Sigma^{-1} = \alpha \mathrm{I}_M + \beta \sum_i \boldsymbol{x}_i \boldsymbol{x}_i^\top \text{ and} \tag{7}$$
$$\mu = \beta \Sigma \sum_i \boldsymbol{x}_i t_i, \tag{8}$$

   and I_M is the M-dimensional identity matrix.

   Hint: Multiply out the log-posterior $\log p(\omega|D, \alpha, \beta) = C + \log p(D|\omega, \beta) + \log p(\omega|\alpha)$, and show that (ignoring constants that do not depend on ω) it is equal to $-\frac{1}{2}(\omega - \mu)^\top \Sigma^{-1} (\omega - \mu)$.

3. **Practical Question:** You can find this week's practical exercise in `Exercise_06.ipynb`