Bayesian methods of Machine learning

Assignment 2

1. (2 points) Consider a data set in which each data point t_n is associated with a weighting factor $r_n > 0$, so that the sum-of-squares error function becomes

$$E_D(\boldsymbol{\theta}) = \frac{1}{2}\sum_{n=1}^{N} r_n \left(t_n - \boldsymbol{\theta}^T \mathbf{x}_n\right)^2 .$$

Find an expression for the solution $\boldsymbol{\theta}^*$ that minimizes this error function. Give two alternative interpretations of the weighted sum-of-squares error function in terms of (i) data dependent noise variance and (ii) replicated data points.

2. (3 points) We have seen that, as the size of a data set increases, the uncertainty associated with the posterior distribution over model parameters decreases. Make use of the Woodbury matrix identity

$$(\mathbf{M} + \mathbf{x}\mathbf{x}^T)^{-1} = \mathbf{M}^{-1} - \frac{(\mathbf{M}^{-1}\mathbf{x})(\mathbf{x}^T\mathbf{M}^{-1})}{1 + \mathbf{x}^T\mathbf{M}^{-1}\mathbf{x}}$$

to show that the uncertainty $\sigma_N^2(\mathbf{x})$ associated with the linear regression function given at Lecture 4, slide 15 satisfies

$$\sigma_{N+1}^2(\mathbf{x}) \leq \sigma_N^2(\mathbf{x}).$$