Machine Learning Worksheet 5

Tomas Ladek, Michael Kratzer
3602673, 3612903
tom.ladek@tum.de, mkratzer@mytum.de

Problem 1

$$\begin{aligned} E_{\mathcal{D}}(w) &= \frac{1}{2}\sum_{n=1}^{N} T_n[\mathbf{W}^T\phi(x_n) - Z_n]^2 \\ &= \frac{1}{2}[T(\Phi W - Z)^T][T(\Phi W - Z)] \end{aligned}$$

with

$$T = \begin{pmatrix} \sqrt{T_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{T_n} \end{pmatrix}$$

Now, finding the optimal W so that this error function is minimal (using the knowledge about the derivative from the slides and the Matrix Cookbook):

$$\begin{aligned} \nabla_W E_{\mathcal{D}}(W) &= \frac{\partial}{\partial W}\frac{1}{2}[T(\Phi W - Z)^T][T(\Phi W - Z)] \\ &= -T^2\Phi^T(\Phi W - Z) \end{aligned}$$

Further, setting this to 0 and solving for W:

$$\begin{aligned} -T^2\Phi^T(\Phi W - Z) &= 0 \\ -T^2\Phi^T\Phi W + T^2\Phi^T Z &= 0 \\ T^2\Phi^T\Phi W &= T^2\Phi^T Z \\ (T^2\Phi^T\Phi)^{-1}(T^2\Phi^T\Phi W) &= (T^2\Phi^T\Phi)^{-1}T^2\Phi^T Z \\ W &= T^{-2}T^2(\Phi^T\Phi)^{-1}\Phi^T Z \\ W &= (\Phi^T\Phi)^{-1}\Phi^T Z \end{aligned}$$

By carefully choosing values of T_n, the error function can be made less sensitive against outliers (that can have a significant impact on the variance of the data noise). Also duplicate data points can be weighted accordingly (e.g. small value of T_i for the data point copies: the error is also made smaller).

Problem 2