CK0146 or TIP8311: Homework 03

Exercise 03.01. You are given a dataset $\mathcal{D}^{(L)}$ (here) consisting of input and output pairs of observations $\{(\mathbf{x}_n^{(L)}, t_n^{(L)})\}_{n=1}^{N_{\text{learning}}}$, such that $\mathbf{x}_n^{(L)} \in \mathbb{R}^D$ with $D = 4$, and $t_n^{(L)} \in \mathbb{R}$. For $\mathcal{D}^{(L)}$, set aside a training set (let $N_{\text{training}} = \dfrac{2}{3} N_{\text{learning}}$) and a validation set ($N_{\text{validation}} = \dfrac{2}{3} N_{\text{learning}}$).

You are requested to build the linear basis function model for regressing the output t onto the space of some $j = 1, \ldots, M$ fixed basis functions $\phi_j(\mathbf{x})$ of the inputs $\mathbf{x} = (x_1, \ldots, x_D)^T$.

Assume that the target variable t is given by a deterministic function $y(\mathbf{x}, \mathbf{w})$ with additive Gaussian noise $\varepsilon \sim \mathcal{N}(0, \beta^{-1})$, so that $t = y(\mathbf{x}, \mathbf{w}) + \varepsilon$ with $t \sim \mathcal{N}(y(\mathbf{x}, \mathbf{w}), \beta^{-1})$, and let $y(\mathbf{x}, \mathbf{w}) = \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x})$ be the regression model, with parameters $\mathbf{w} = (w_0, \ldots, w_{M-1})^T$ and $\boldsymbol{\phi} = (\phi_0, \ldots, \phi_{M-1})^T$ with $\phi_0(\mathbf{x}) = 1$. Also, assume that the output observations $\{t_n\}$ are drawn independently from distribution $\mathcal{N}(y(\mathbf{x}, \mathbf{w}), \beta^{-1})$ with β unknown.

Determine an optimal set of model parameters $\mathbf{w}$ under the two following modelling setups:

- **A**: For a varying number $j = 1, 2, \ldots, M_{\text{max}}$ of (fixed) basis, you must estimate the model parameters using the training set (to get, say, $\hat{\mathbf{w}}_{\text{training}}(j)$) and assess the performance of the resulting model on the validation set (to get, say, $E(\mathbf{w}^*_{\text{training}}(j))$). Based on the validation results, you must select the optimal number $\hat{M}$ of basis functions (as the one that corresponds to the smallest error) and estimate the parameters of the final model (that is, $\hat{\mathbf{w}}_{\text{training+validation}}(\hat{M})$) on the rejoined training and validation sets.

- **B**: For a fixed number of (fixed) basis functions (for instance, $\hat{M}$ from above), you must regularise the error function by penalising the L_2 norm of the parameter vector. For a varying regularisation parameter $\lambda \in (\lambda_{\text{min}}, \lambda_{\text{max}}) \subseteq \{0, \mathbb{R}^+\}$, you must estimate the model parameters using the training set (to get, say, $\hat{\mathbf{w}}_{\text{training}}(\lambda)$) and assess the performance of the resulting model on the validation set (to get, say, $E(\mathbf{w}^*_{\text{training}}(\lambda))$). Based on the validation results, you must select the value $\hat{\lambda}$ of the regularisation parameter (as the one that corresponds to the smallest error) and estimate the parameters of the final model (that is, $\hat{\mathbf{w}}_{\text{training+validation}}(\hat{\lambda})$) on the rejoined training and validation sets.

Once the models have been defined ($\hat{\mathbf{w}}_{\text{training+validation}}(\hat{M})$ and $\hat{\mathbf{w}}_{\text{training+validation}}(\hat{\lambda})$ have been determined), you must asses their performance on the independent test set $\mathcal{D}^{(T)}$ (here).

[*Note*] To assess the model performances, you must use the root-mean-square (RMS) error, at all stages (training, validation, test). You are free to choose your favourite family of basis functions (that is, Gaussians, sigmoids, etc.), its internal parameters and its arrangement.

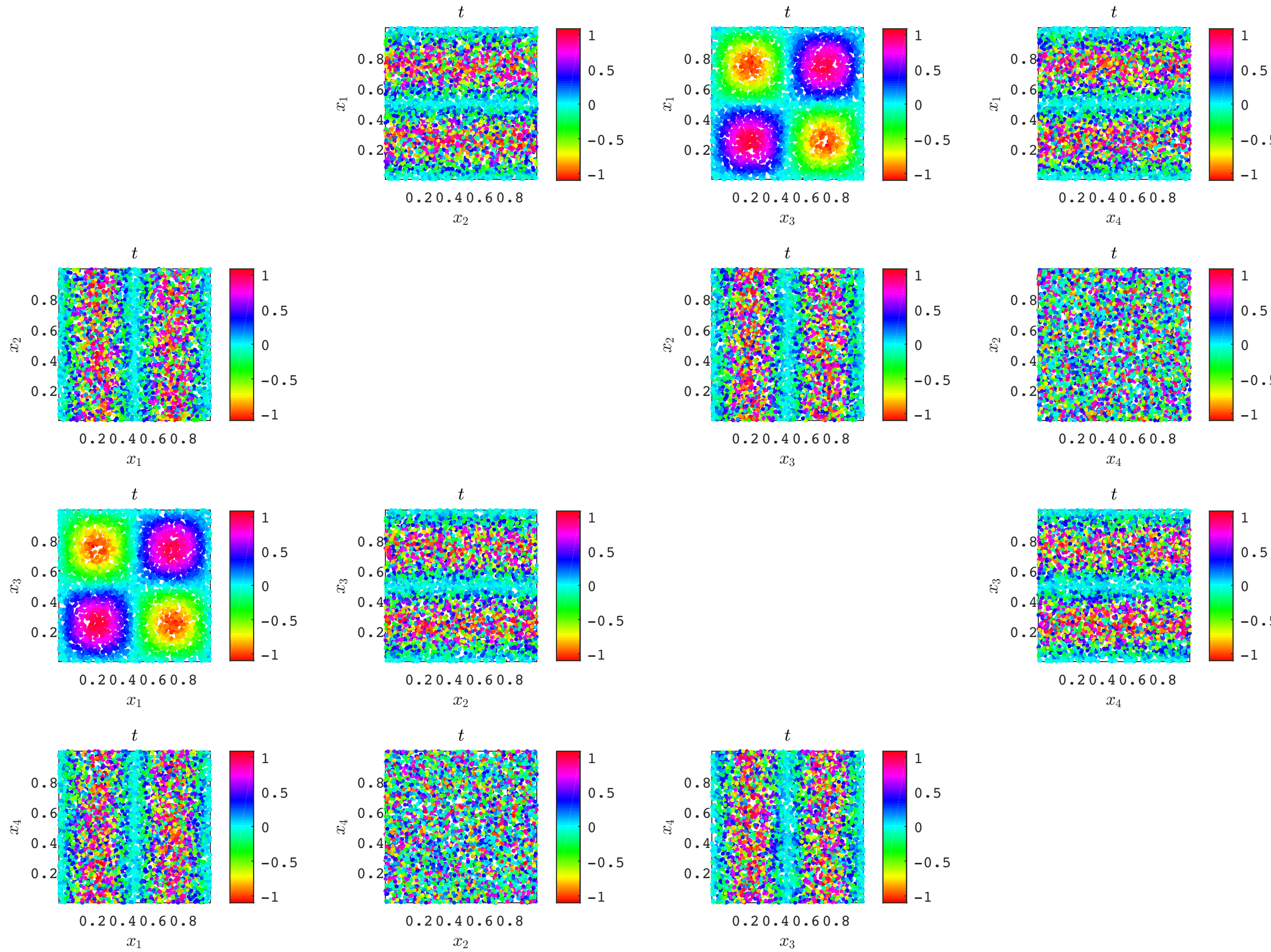

Figure 1: Pairwise scatterplots of the $4D$ input vectors $\mathbf{x} = (x_1, \dots, x_4)^T$, dyed using the corresponding value of the outputs t.