Nonlinear Basis Functions

a) Monomials in $\mathbb{R}^N$ of highest degree k:

$$\phi_i(\underline{x}) \in \{1, x_1, x_2, \dots, x_N, x_1^2, x_1 x_2, x_1 x_3, \dots, x_1 x_N, \dots x_N^2, \dots, x_1 x_N^{k-1}, x_1^k, \dots, x_N^k\}$$

$$= \left\{ \prod_{j=1}^{N} x_j^{a_j} \; : \; a_j \in \mathbb{N}_0 \text{ with } 0 \le \sum_{j=1}^{N} a_j \le k \right\}, \quad i = 1, \dots, d$$

Homework: $k = 9 \; (\Rightarrow d = 55), \; N = 2$

b) Sphering:

$\phi_i(\underline{x})$ can grow fast for large $\|\underline{x}\|$ $\Rightarrow$ whitening (sphering) which includes standardization

$\rightarrow$ scale and rotate centered data s.t. resulting ("sphered") data is uncorrelated with standard deviation 1 in each variable

$\forall$ data pts. α:

$$\underline{x}^{(\alpha)}_{\text{sphered}} = \underline{\Lambda}^{-\frac{1}{2}} \underline{E}^T \underline{x}^{(\alpha)}_{\text{centered}}$$

where $\underline{x}^{(\alpha)}_{\text{centered}} = \underline{x}^{(\alpha)} - \underline{\bar{x}} \quad \left(\underline{\bar{x}} = \frac{1}{p} \sum_{\alpha=1}^{p} \underline{x}^{(\alpha)} \right)$

$\underline{\Lambda} := \operatorname{diag}(\lambda_1, \dots, \lambda_N)$ eigenvalue matrix, $\underline{E} = (\underline{e}_1, \dots, \underline{e}_N)$ eigenvector matrix of the covariance matrix (of the data) $\underline{C}$, i.e., $\underline{C}\,\underline{e}_i = \lambda_i \underline{e}_i$.

$\underline{C}$ is given by $C_{ij} = \frac{1}{p} \sum_{\alpha=1}^{p} x^{(\alpha)}_{\text{centered},i} x^{(\alpha)}_{\text{centered},j}$ (i.e. $\underline{C} = \frac{1}{p} \underline{X}_c \underline{X}_c^T$ for $\underline{X}_c = \left(\underline{x}^{(1)}_{\text{centered}}, \dots, \underline{x}^{(p)}_{\text{centered}} \right)$)

$\Rightarrow \frac{1}{p} \sum_{\alpha=1}^{p} \underline{x}^{(\alpha)}_{\text{sphered}} = \underline{0}$ mean (zero $\hat{=}$ centered)

and $\frac{1}{p} \sum_{\alpha=1}^{p} \underline{x}^{(\alpha)\,2}_{\text{sphered}} = 1$ variance (unit std. dev.)

and $\frac{1}{p} \sum_{\alpha=1}^{p} x^{(\alpha)}_{\text{sphered},i} \, x^{(\alpha)}_{\text{sphered},j} = 0$ correlation (uncorrelated variables)

c) Analytical solution for linear neuron with basis functions, quadratic cost, regularization (L_2)

Data Set $\underline{\phi}^{(\alpha)} := \underline{\phi}(\underline{x}^{(\alpha)}) := \begin{pmatrix} \phi_1(\underline{x}^{(\alpha)}) \\ \vdots \\ \phi_d(\underline{x}^{(\alpha)}) \end{pmatrix}, \; \alpha = 1, \dots, p$ [+ labels $y_T^{(\alpha)}$]

$\Rightarrow$ From previous results: $\underline{w}^* = \left(\underline{\Phi}\, \underline{\Phi}^T + \underbrace{\lambda \underline{I}}_{=0 \text{ w/o regularization, i.e. } \lambda = 0} \right)^{-1} \underline{\Phi}\, \underline{y}^T$

where $\underline{\Phi} = \left(\underline{\phi}^{(1)}, \dots, \underline{\phi}^{(p)} \right) \in \mathbb{R}^{d,p}$

$y(\underline{x}; \underline{w}) = \underline{w}^T \underline{\phi}(\underline{x}) = \sum_{i=1}^{d} w_i \phi_i(\underline{x})$