

# Linear Regression

This assignment shows how we can extend ordinary least squares regression, which uses the hypothesis class of linear regression functions, to non-linear regression functions modeled using polynomial basis functions and radial basis functions. The function we want to fit is  $y_{\text{true}} = f_{\text{true}}(x) = 6(\sin(x + 2) + \sin(2x + 4))$ . This is a **univariate function** as it has only one input variable. First, we generate synthetic input (data)  $x_i$  by sampling  $n = 750$  points from a uniform distribution on the interval  $[-7.5, 7.5]$ .

## 1. Regression with Polynomial Basis Functions

This problem extends **ordinary least squares regression**, which uses the hypothesis class of *linear regression functions*, to *non-linear regression functions* modeled using **polynomial basis functions**. In order to learn nonlinear models using linear regression, we have to explicitly **transform the data** into a higher-dimensional space. The nonlinear hypothesis class we will consider is the set of  $d$ -degree polynomials of the form  $f(x) = w_0 + w_1 x + w_2 x^2 + \dots + w_d x^d$  or **a linear combination of polynomial basis function**:

$$f(x) = [w_0, w_1, w_2, \dots, w_d]^T \begin{bmatrix} 1 \\ x \\ x^2 \\ \vdots \\ x^d \end{bmatrix}.$$

The monomials  $\{1, x, x^2, \dots, x^d\}$  are called **basis functions**, and each basis function  $x^k$  has a corresponding weight  $w_k$  associated with it, for all  $k = 1, \dots, d$ . We transform each univariate data point  $x_i$  into a multivariate ( $d$ -dimensional) data point via  $\phi(x_i) \rightarrow [1, x_i, x_i^2, \dots, x_i^d]$ . When this transformation is applied to every data point, it produces the **Vandermonde matrix**:

$$\Phi = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^d \\ 1 & x_2 & x_2^2 & \dots & x_2^d \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^d \end{bmatrix}.$$

## 2. Regression with Radial Basis Functions

In the previous case, we considered a nonlinear extension to linear regression using a linear combination of polynomial basis functions, where each basis function was introduced as a feature  $\phi(x) = x^k$ . Now, we consider Gaussian radial basis functions of the form:

$$\phi(\mathbf{x}) = e^{-\gamma(x-\mu)^2},$$

whose shape is defined by its center  $\mu$  and its width  $\gamma > 0$ . In the case of polynomial basis regression, the user's choice of the dimension  $d$  determined the transformation and the model. For radial basis regression, we have to contend with deciding how many radial basis functions we should have, and what their center and width parameters should be. For simplicity, let's assume that  $\gamma = 0.1$  is fixed. Instead of trying to identify the number of radial basis functions or their centers, we can treat **each data point as the center of a radial basis function**, which means that the model will be:

$$f(x) = [w_1, w_2, w_3 \dots, w_n]^T \begin{bmatrix} e^{-\gamma(x-x_1)^2} \\ e^{-\gamma(x-x_2)^2} \\ e^{-\gamma(x-x_2)^2} \\ \dots \\ e^{-\gamma(x-x_n)^2} \end{bmatrix}.$$

This transformation uses radial basis functions centered around data points  $e^{-\gamma(x-x_i)^2}$  and each basis function has a corresponding weight  $w_i$  associated with it, for all  $i = 1, \dots, n$ . We transform each univariate data point  $x_j$  into into a multivariate ( $n$ -dimensional) data point via  $\phi(x_j) \rightarrow [\dots, e^{-\gamma(x_j-x_i)^2}, \dots]$ . When this transformation is applied to every data point, it produces the **radial-basis kernel**:

$$\Phi = \begin{bmatrix} 1 & e^{-\gamma(x_1-x_2)^2} & e^{-\gamma(x_1-x_3)^2} & \dots & e^{-\gamma(x_1-x_n)^2} \\ e^{-\gamma(x_2-x_1)^2} & 1 & e^{-\gamma(x_2-x_3)^2} & \dots & e^{-\gamma(x_2-x_n)^2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e^{-\gamma(x_n-x_1)^2} & e^{-\gamma(x_n-x_2)^2} & e^{-\gamma(x_n-x_3)^2} & \dots & 1 \end{bmatrix}.$$