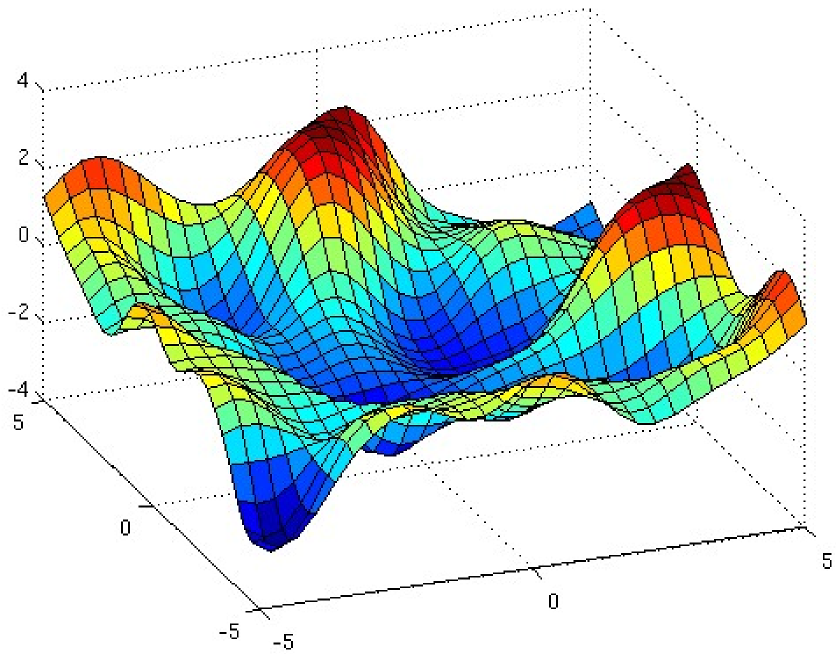
Regression models

Dr. Pritam Anand.
Assistant Professor,
DA-IICT, Gandhinagar.

Polynomial Basis Functions

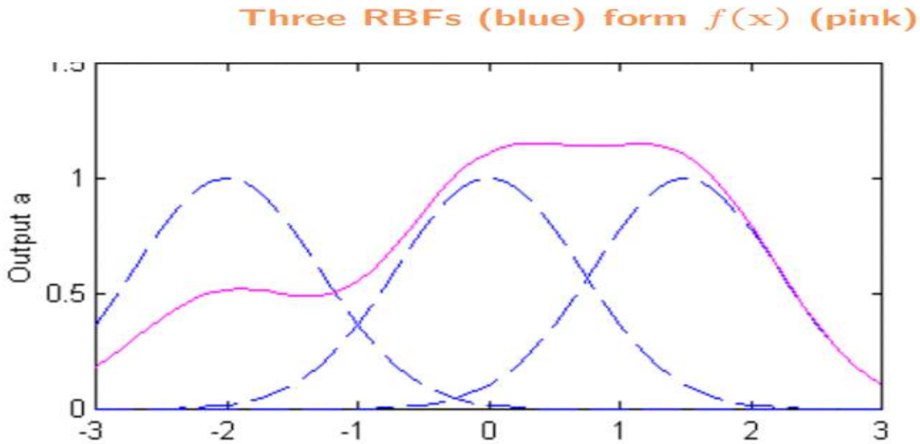
$$f(\boldsymbol{x}) = \boldsymbol{W}^T\phi(\boldsymbol{x}), \quad w \in R^{\frac{(m+n)!}{m!+n!}}$$

$$\phi(\boldsymbol{x}) = \begin{bmatrix} \phi_M(\boldsymbol{x}) \\ \phi_{M-1}(\boldsymbol{x}) \\ - \\ - \\ \phi_1(\boldsymbol{x}) \\ \phi_0(\boldsymbol{x}) \end{bmatrix} = \begin{bmatrix} {x_1}^m \\ {x_2}^m \\ - \\ - \\ x_n \\ 1 \end{bmatrix}$$



Gaussian Basis functions.

$$f(\boldsymbol{x}) = \boldsymbol{W}^T \phi(\boldsymbol{x})$$



Gaussian function in R

$$\phi(\boldsymbol{x}) = \begin{bmatrix} \phi_M(\boldsymbol{x}) \\ \phi_{M\,-\,1}(\boldsymbol{x}) \\ - \\ - \\ \phi_1(\boldsymbol{x}) \\ \phi_0(\boldsymbol{x}) \end{bmatrix}, where\ \phi_j(\boldsymbol{x}) = \exp\left(-\frac{1}{2s_j}\,||\boldsymbol{x} - \mathbf{c_j}||^2\right)$$

Sigmoidal Basis functions

$$\min_{f} \sum_{i=1}^{2} (y_i - f(x_i))^2 + \lambda \| f$$

$$f(\boldsymbol{x}) = \boldsymbol{W}^T \phi(\boldsymbol{x})$$

$$\phi(\boldsymbol{x}) = \begin{bmatrix} \phi_M(\boldsymbol{x}) \\ \phi_{M-1}(\boldsymbol{x}) \\ - \\ - \\ \phi_1(\boldsymbol{x}) \\ \phi_0(\boldsymbol{x}) \end{bmatrix},$$

$$where\ \phi_j(\boldsymbol{x}) = \sigma\ (\boldsymbol{x}, \boldsymbol{c}_j, bj) = \frac{e^{\boldsymbol{c}_j{}^T\boldsymbol{x}+bj}}{1 + e^{\boldsymbol{c}_j{}^T\boldsymbol{x}+bj}}, cj \in Rn, bj \in R.$$

Feature Map

$$\begin{bmatrix} x_1 \\ x_2 \\ \\ x_n \end{bmatrix}$$

$$f(\boldsymbol{x}) = \boldsymbol{W}^T \phi(\boldsymbol{x}) + \mathrm{b}$$

$$\phi(\boldsymbol{x}) = \begin{bmatrix} \phi_M(\boldsymbol{x}) \\ \phi_{M-1}(\boldsymbol{x}) \\ - \\ - \\ \phi_1(\boldsymbol{x}) \\ 1 \end{bmatrix}$$

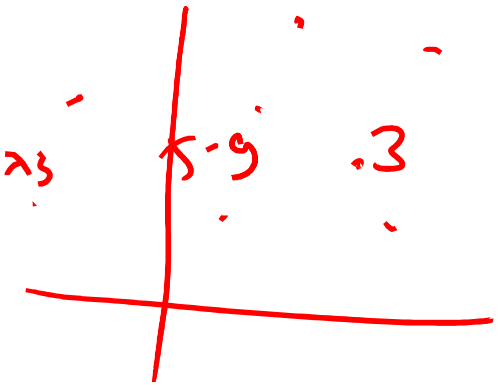
Basis Functions and Feature Map

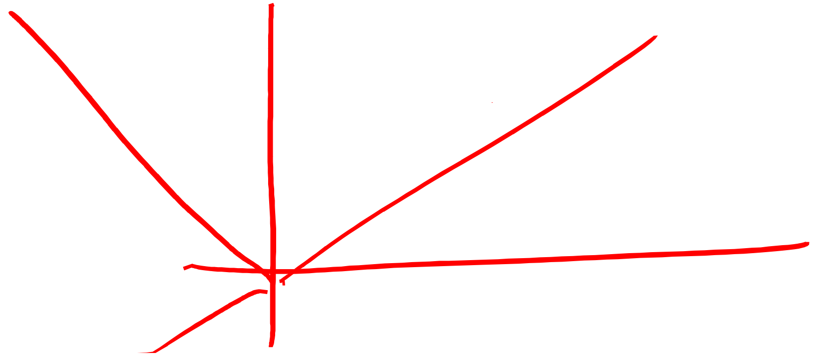
x1	x2	Y
5.9	3	49.81
6.9	3.1	63.23
6.6	2.9	57.98
4.6	3.2	37.41
6	2.2	46.85

$$\phi(x) = \begin{bmatrix} \phi_1(x) \\ \phi_2(x) \\ \phi_3(x) \\ \phi_4(x) \\ \phi_5(x) \end{bmatrix} = \begin{bmatrix} x_1^2 \\ x_2^2 \\ x_1 x_2 \\ x_1 \\ x_2 \end{bmatrix}$$

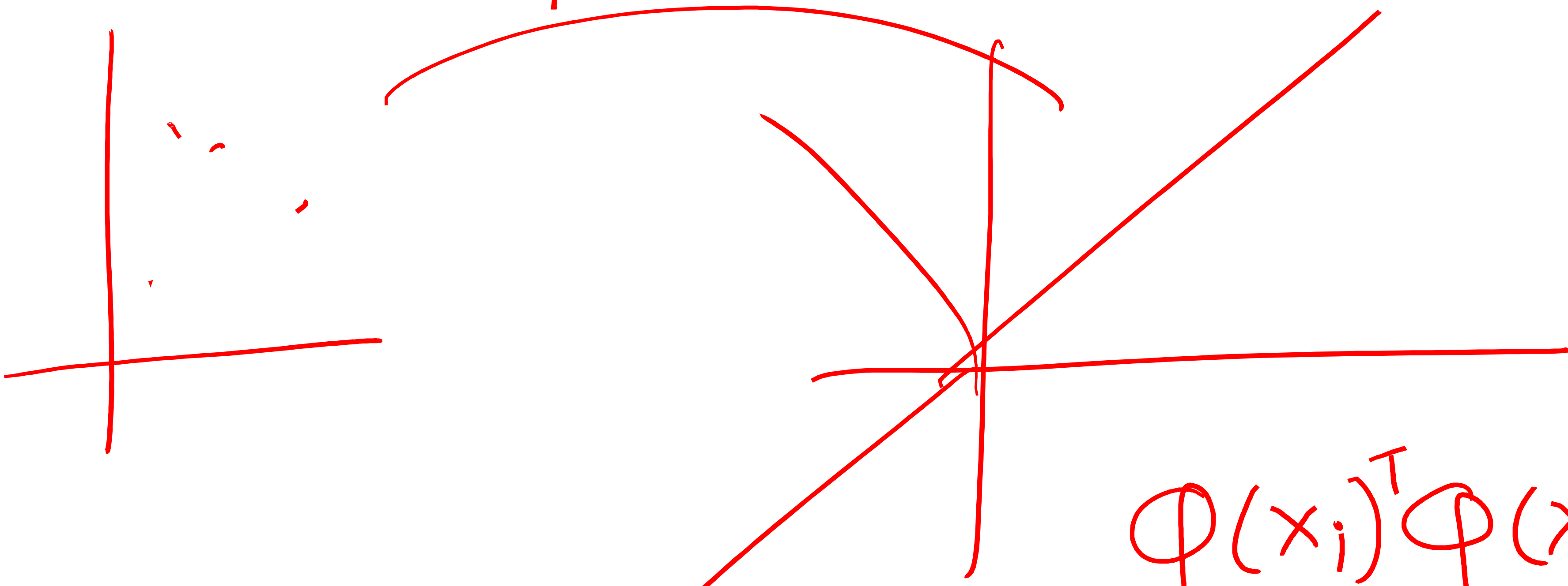
$$w^T \phi(x) + b$$

$$(5.9)^2 \quad 3^2 \quad 5.9 \times 3 \quad 5.9 \quad 3 \quad 1$$





$X \xrightarrow{\varphi} \mathcal{H}$ $\quad w =$



$\varphi(x_i)^T \varphi(x_j)$

$K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$

$$f(x) = w^T \varphi(x) + b \qquad w = \sum_{i=1}^{l} u_i \varphi(x_i)$$

$$= \left(\sum_{i=1}^{l} u_i \varphi(x_i) \right)^T \varphi(x) + b$$

$$= u_1 \varphi(x_1) + u_2 \varphi(x_2) + \cdots + u_l \varphi(x_l)$$

$$= \sum_{i=1}^{l} u_i \left(\varphi(x_i)^T \varphi(x) \right) + b$$

$$(x + y)^T z = x^T z + y^T z$$

$$= \sum_{i=1}^{l} K(x_i, x) u_i + b$$

$$K = \begin{bmatrix} K(x_1, x_1) & K(x_1, x_2) & \cdots & K(x_1, x_l) \\ K(x_2, x_1) & K(x_2, x_2) & & K(x_2, x_l) \end{bmatrix}$$

$$K = U^T D U$$

Feature Map

$$\sum_{i=1}^{l} K(x_i, x) y_i + b$$

$$K(x_1, x) y_1 + K(x_2, x) y_2 + \cdots + K(x_l, x) y_l + b$$

$$f(\boldsymbol{x}) = \boldsymbol{W}^T \phi(\boldsymbol{x}) + b$$

$$\phi(\boldsymbol{x}) = \begin{bmatrix} \phi_M(\boldsymbol{x}) \\ \phi_{M-1}(\boldsymbol{x}) \\ - \\ - \\ \phi_1(\boldsymbol{x}) \\ 1 \end{bmatrix}$$

Use this feature map

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_l \end{bmatrix}^T \begin{bmatrix} K(\boldsymbol{x}_1, \boldsymbol{x}_1) \\ K(\boldsymbol{x}_1, \boldsymbol{x}_2) \\ - \\ - \\ K(\boldsymbol{x}_l, \boldsymbol{x}_l) \\ \cancel{1} \end{bmatrix} + b$$

$$f(x) = \sum_{i=1}^{l} K(x_i, x)\, u_i + b$$

$$\min_{u,\,b} \sum_{i=1}^{l} \left(y_i - \left(\sum_{i=1}^{l} K(x_i, x)\, u_i + b \right) \right)^2 + \lambda\, u^T u$$

$$u = \begin{bmatrix} u_1 \\ u_2 \\ - \\ u_l \end{bmatrix} \qquad \begin{bmatrix} u_1 \\ u_2 \\ - \\ u_l \\ b \end{bmatrix} = \left(A^T A + \lambda I \right)^{-1} A^T y$$

$$A = \begin{bmatrix} K(x_1, x_1) & K(x_1, x_2) \cdots & K(x_1, x_l) & 1 \\ & & & \\ K(x_l, x_1) & K(x_l, x_2) & K(x_l, x_l) & 1 \end{bmatrix}$$

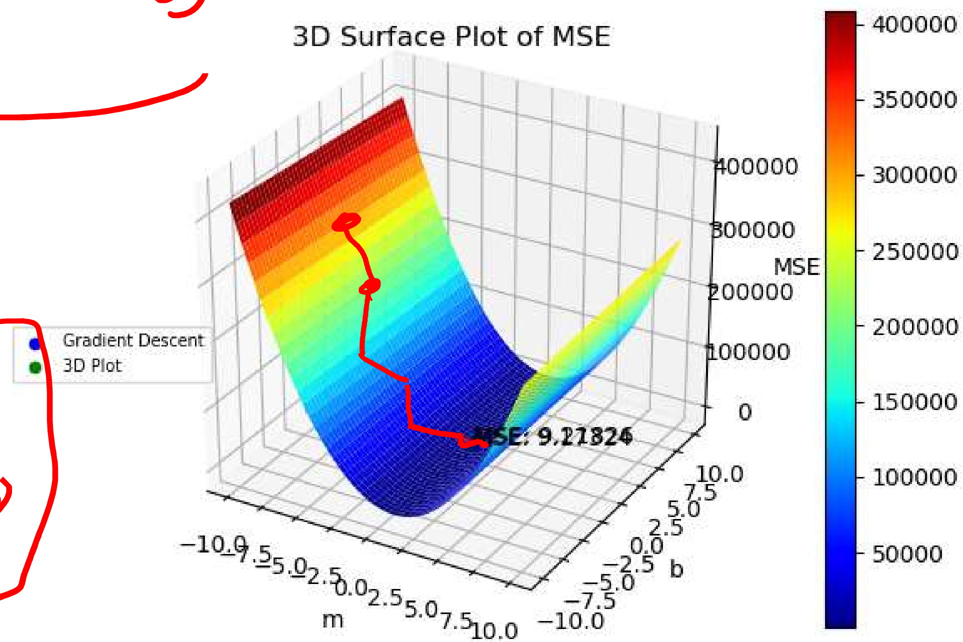
Least Square Regression model with gradient descent

$$\min_{w,b} \underbrace{\frac{\lambda w^T w}{2} + \sum_{i=1}^{2} \left(y_i - (w^T x_i + b)\right)^2}_{} = F(w,b)$$

$\rightarrow \lambda w$

$\begin{bmatrix} w^0 \\ b^0 \end{bmatrix}$

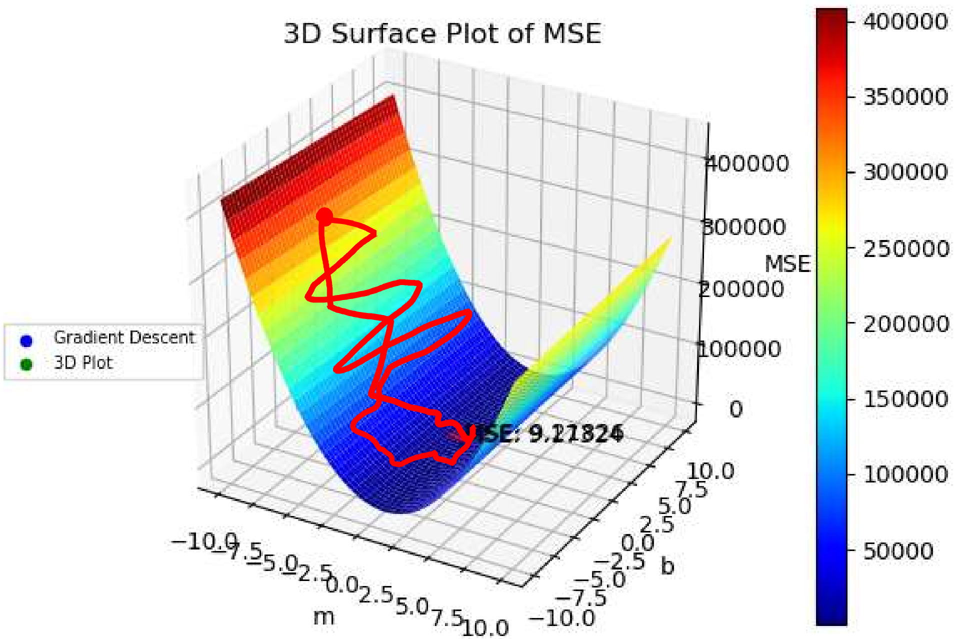
$$\begin{bmatrix} w^{(1)} \\ b^{(1)} \end{bmatrix} = \begin{bmatrix} w^{(0)} \\ b^{(0)} \end{bmatrix} - \eta \begin{bmatrix} \nabla_w F(w,b) \\ \nabla_b F(w,b) \end{bmatrix}$$



Compute the gradient

$$\sum_{k=1}^{L} \left(y_k - (w^T x_k + b) \right)^2$$

$$\begin{bmatrix} -2(y_k - (w^T x_k + b))\, x_k \\ -2(y_k - (w^T x_k + b))\, 1 \end{bmatrix}$$



Gradient Descent Method

$$\frac{\lambda}{2} w^T w + \sum_{i=1}^{l} \left(y_i - (w^T \phi(x_i) + b) \right)^2$$

λw

First Let us consider

$$\mathrm{J}(\mathrm{u}, \mathrm{b}, \mathrm{x_k}, \mathrm{y_k}) \;=\; (y_k \,-\, (w^T \phi(xk) + b\,)^2$$

Least Square Kernel Regression

First Let us consider

$$J(\mathrm{w}, \mathrm{b}, \mathrm{x_k}, \mathrm{y_k}) \;=\; (y_k \;-(w^{\,T}\phi(xk) + b))^2$$

$$\frac{\partial\, J\,(\mathrm{w,b,x_k,y_k})}{\partial w} = -\,2\,(y_k \;-(w^{\,T}\phi(xk) + b))\phi(x_k)$$

$$\frac{\partial\, J\,(\mathrm{w,b,x_k,y_k})}{\partial b} = -\,2\,(y_k \;-(w^{\,T}\phi(xk) + b))$$

Gradient Descent Least Square Kernel Regression

Algorithm:- Gradient descent method

Initialize $x^0 = w^{start} \in RM$ and $b \in R$

Repeat

$$w^{(j+1)} := w^{(j)} - \eta_k \left(\lambda w + \sum_{i=1}^{l} \frac{\partial J(w,b,x_k,y_k)}{\partial w} \right).$$

$$b^{(j+1)} := b^{(j)} - \eta_k \left(\sum_{i=1}^{l} \frac{\partial J(w,b,x_k,y_k)}{\partial b} \right)$$

$$\text{Until } \left\| \begin{bmatrix} \lambda w + \sum_{i=1}^{l} \frac{\partial J(w,b,x_k,y_k)}{\partial w} \\ \sum_{i=1}^{l} \frac{\partial J(w,b,x_k,y_k)}{\partial b} \end{bmatrix} \right\| \leq \varepsilon$$

Stochastic Gradient Descent Least Square Regression

Algorithm:- Stochastic Gradient descent method

Initialize $w^0 = w^{start} \in RM$ and $b^0 \in R$

Repeat

Randomly select subset B from Training set T.

$$w^{(j+1)} := w^{(j)} - \eta_k \left(\lambda w + \sum_{(x_k, y_k) \in B} \frac{\partial J(w^{(j)}, b^{(j)}, x_k, y_k)}{\partial w} \right).$$

$$b^{(j+1)} := b^{(j)} - \eta_k \left(\sum_{(x_k, y_k) \in B} \frac{\partial J(w^{(j)}, b^{(j)}, x_k, y_k)}{\partial b} \right).$$

Until $\left\| \begin{bmatrix} w^{(j+1)} \\ b^{(j+1)} \end{bmatrix} - \begin{bmatrix} w^{(j)} \\ b^{(j)} \end{bmatrix} \right\| \leq \varepsilon$