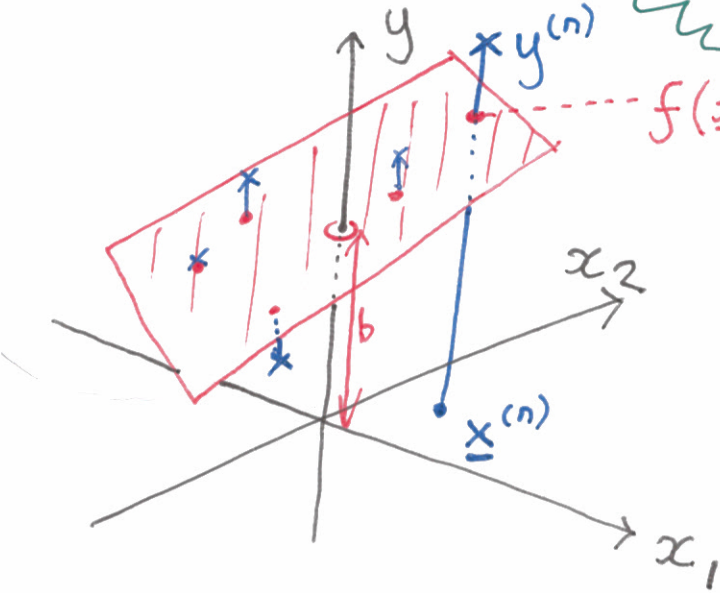
# MLRR Lecture 3

{tinyurl.com/edmlpr}




$$f(\underline{x}^{(n)}; \underline{w}, b) = \underline{w}^T \underline{x}^{(n)} + b = \underline{v}^T \underline{\phi}(\underline{x}^{(n)})$$

$$\underline{v} = \begin{bmatrix} b \\ \underline{w} \end{bmatrix}, \quad \phi(\underline{x}) = \begin{bmatrix} 1 \\ \underline{x} \end{bmatrix}$$

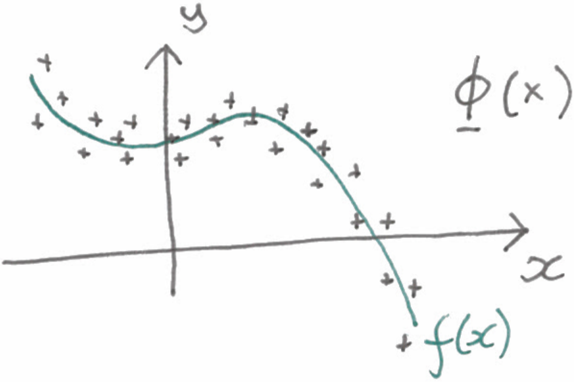
$\underline{f} = \Phi \, \underline{v}$, choose $\underline{v}$ to minimize

$$(\underline{y} - \underline{f})^T (\underline{y} - \underline{f})$$

$n^{th}$ row $\underline{\phi}(\underline{x}^{(n)})^T$

"$\underline{v} = \Phi \backslash \underline{y}$"

$f_n = f(\underline{x}^{(n)}; \underline{v})$



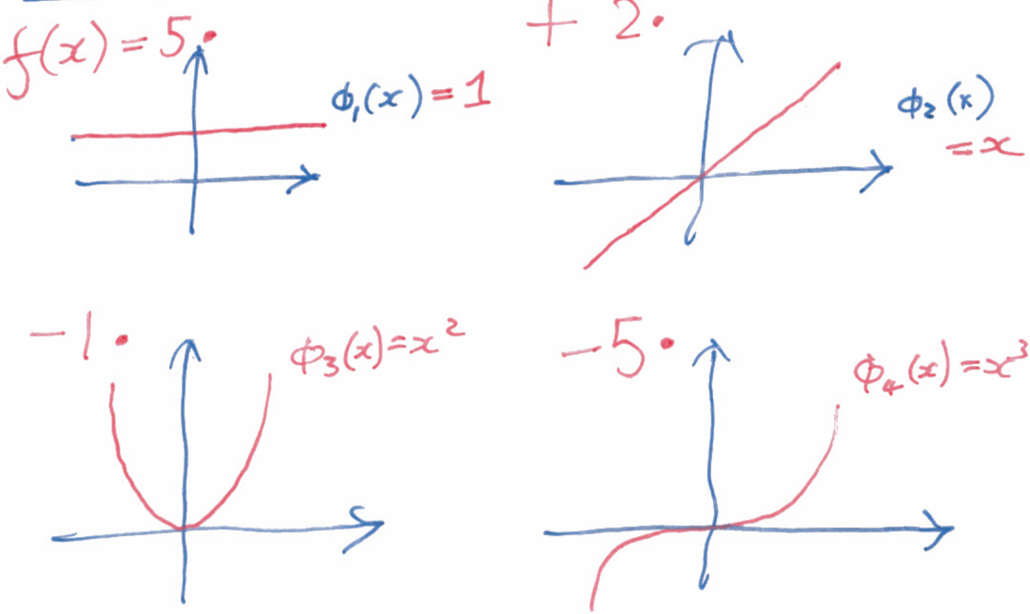

$$\underline{\phi}(x) = [1 \;\; x \;\; x^2 \;\; x^3]^T$$

Fit

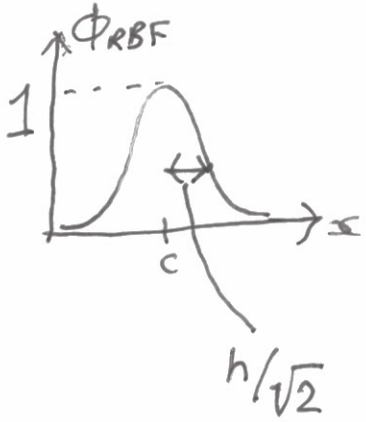
$$\underline{y} \approx \underline{f} = \Phi \underline{w}$$

$$f(x) = w_1 + w_2 x + w_3 x^2 + w_4 x^3$$

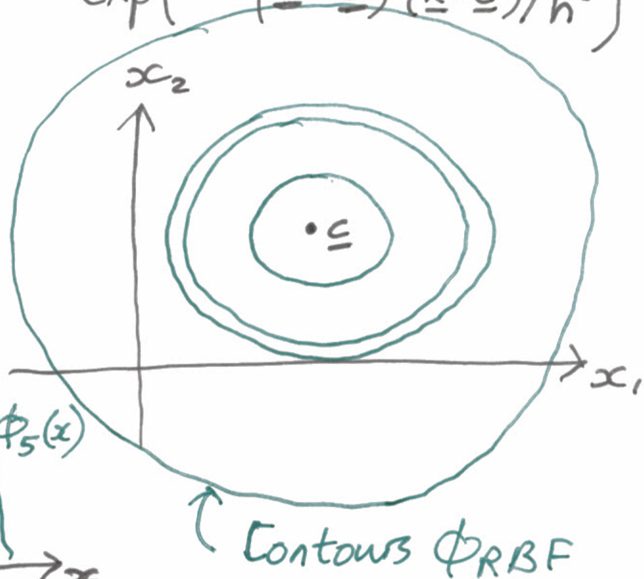
## Basis functions

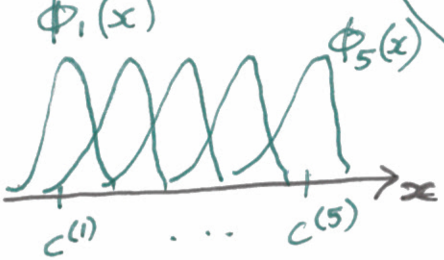


# Radial Basis Function (RBFs)

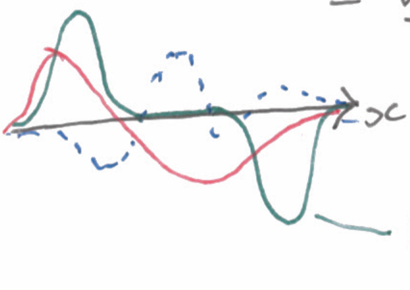



$$\phi_{RBF}(\underline{x}\,;\,\underline{c}\,,h)$$
$$= \exp\left(-(\underline{x}-\underline{c})^T(\underline{x}-\underline{c})/h^2\right)$$




$$f(x) = \sum_{k=1}^{5} w_k \phi_k(x)$$
$$= \underline{w}^T \phi(\underline{x})$$

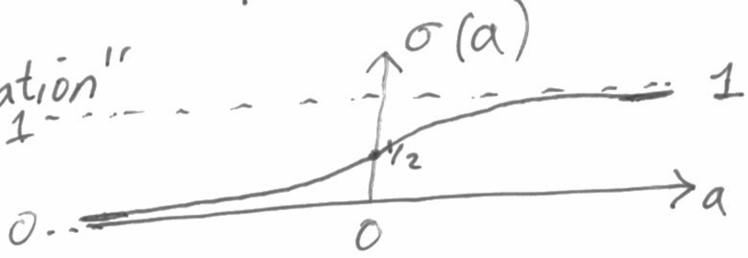



$\underline{w} = [1\ 0\ 0\ 0\ -1]^T$

# Logistic - Sigmoid function

$$\sigma(a) = \frac{1}{1+e^{-a}}$$

"activation"




## Basis $f^{\underline{n}}$

$$\phi_\sigma(\underline{x}; \underline{v}, b) = \sigma\left(\underline{v}^T \underline{x} + b\right)$$

To do yourself: 2D contour plot

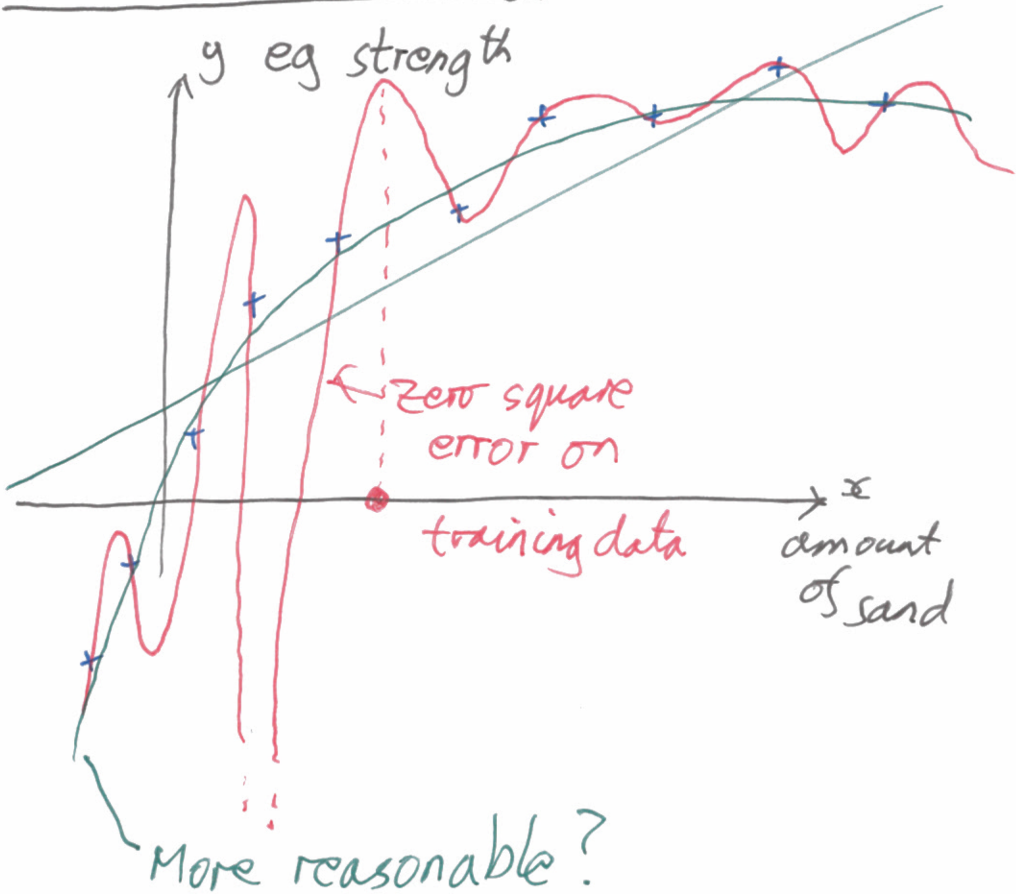
# High-dimensional Polynomials

$$\underline{\phi}(\underline{x}) = [1 \quad x_1 \quad x_2 \quad x_3 \quad \ldots$$
$$x_1^2 \quad x_2^2 \quad x_3^2 \quad \ldots$$
$$x_1 x_2 \quad x_1 x_3 \quad x_2 x_3 \quad \ldots$$
$$x_1^3 \quad x_1 x_2 x_3 \quad x_1 x_2^2 \quad \ldots$$
$$\ldots]$$

Monomial

(Binary vectors can work well)

# Under / Over Fitting




## L2 Regularization

Discourage extreme fits

$\underline{w}^T\underline{w} = \|\underline{w}\|^2$ should be small

Cost function, which we minimize

$$C_\lambda(\underline{w}) = (\underline{y} - \underbrace{\underline{\Phi}\,\underline{w}}_{\underline{f}})^T(\underline{y} - \underline{\Phi}\,\underline{w}) + \lambda\,\underline{w}^T\underline{w}$$

$$\lambda \in [0, \infty]$$

$$\underline{y}' = \begin{bmatrix} \underline{y} \\ \underline{0} \end{bmatrix} \qquad \underline{\Phi}' = \begin{bmatrix} \underline{\Phi} \\ \sqrt{\lambda}\,\mathbb{I} \end{bmatrix}$$

$$C_\lambda(\underline{w}) = (\underline{y}' - \underline{\Phi}'\underline{w})^T(\underline{y}' - \underline{\Phi}'\underline{w})$$

Fit $\hat{\underline{w}} = \underline{\Phi}' \backslash \underline{y}'$

means fitted