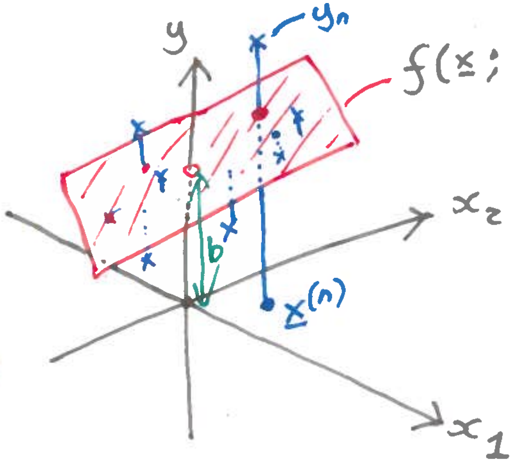
MLPR Lecture 3

12:10pm
and again at 1:10pm

tinyurl.com/edmlpr



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

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

$$\underline{f} = \underline{\Phi}\, \underline{v},$$

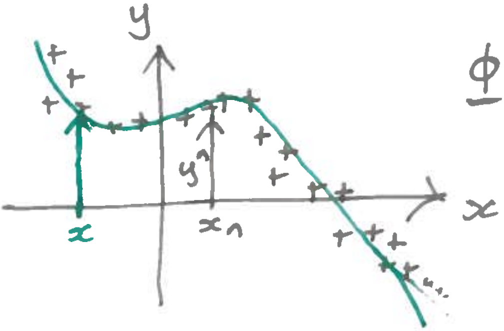
n^{th} row $\underline{\phi}(\underline{x}^{(n)})^T$

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

choose $\underline{v}$ to minimize

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

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



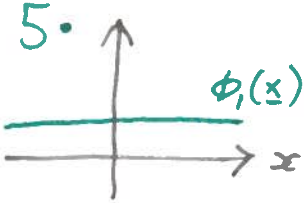
$$\underline{\phi}(\underline{x}) = [1 \;\; x \;\; x^2 \;\; x^3]^T$$

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

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

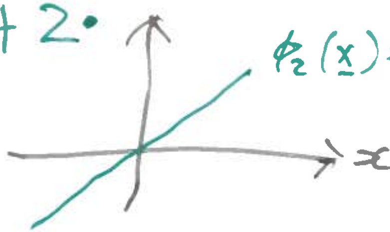
Basis Functions

$f(x) = 5 \cdot$

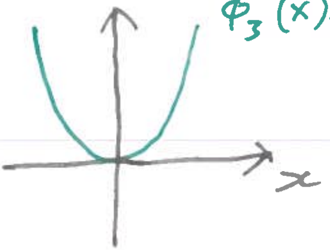


$\phi_1(\underline{x}) = 1$

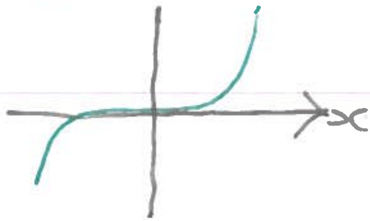
$+ 2 \cdot$ $\phi_2(\underline{x}) = x$



$-1 \cdot$ $\phi_3(x) = x^2$

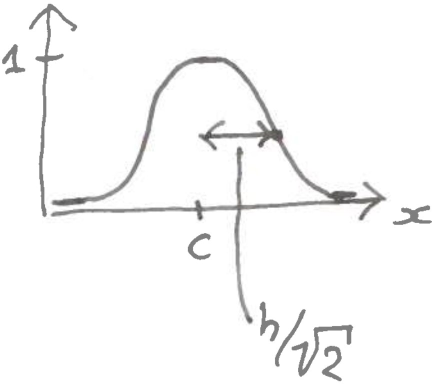


$-5 \cdot$ $\phi_4(x) = x^3$



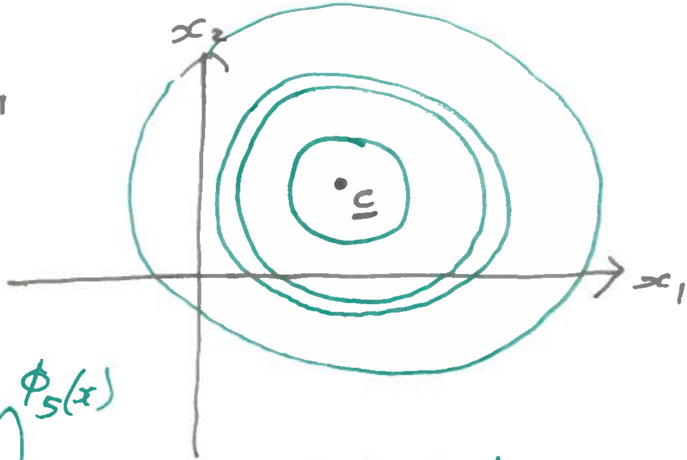
We could choose any other basis $f^{\underline{n}}$'s

Radial Basis Functions (RBFs)

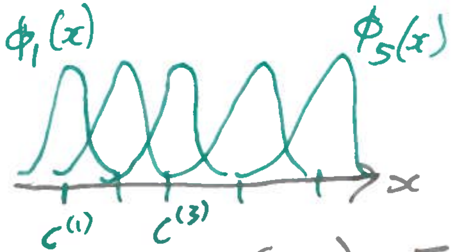


$$\phi_{RBF}(\underline{x};\underline{c},h)$$

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

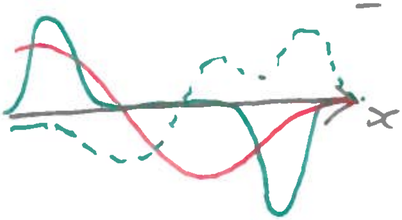


Contours ϕ_{RBF}



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

$$= \underline{w}^T \underline{\phi}(\underline{x})$$

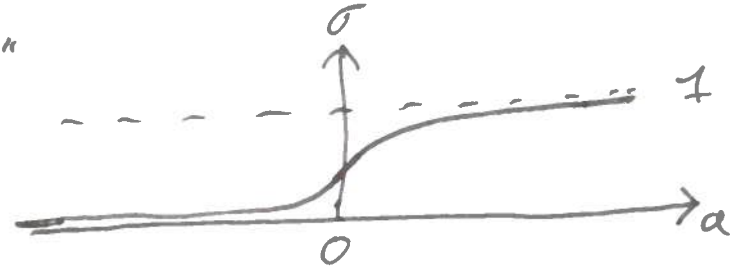


$f\underline{n}$'s for different $\underline{w}$'s

Logistic - Sigmoid function

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

"activation"



Basis f^n

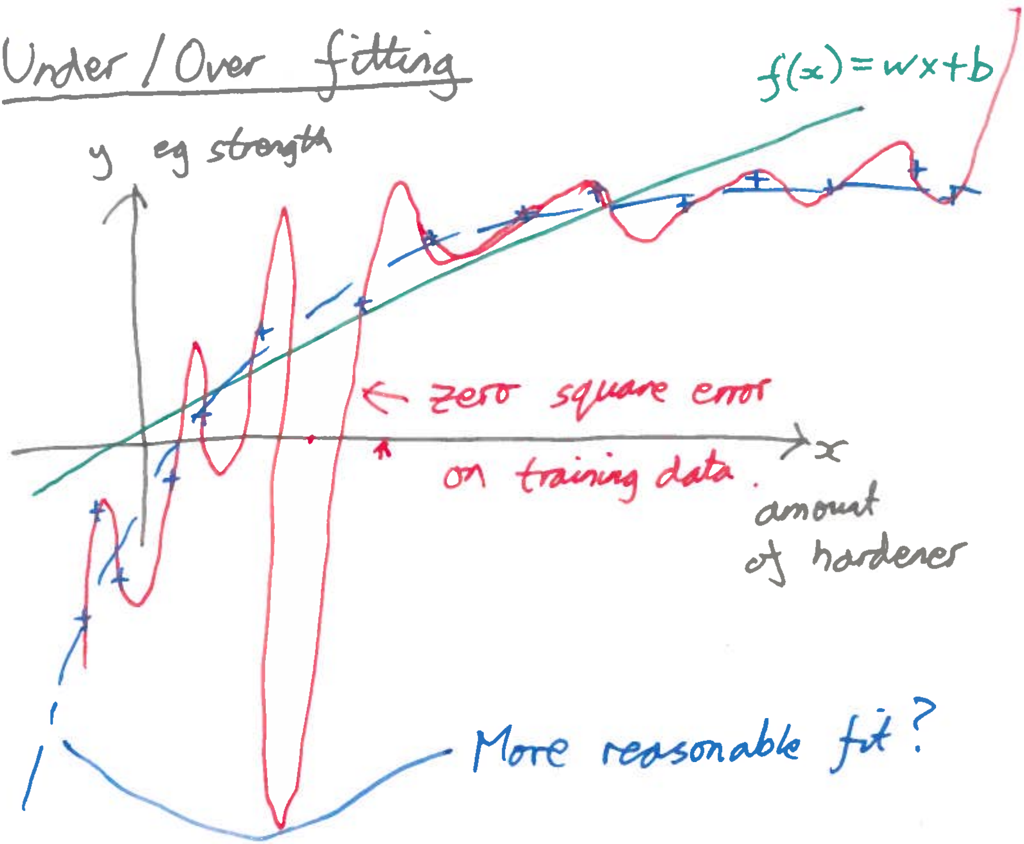
$$\phi_\sigma(\underline{x}; \underline{v}, b) = \sigma(\underline{v}^T \underline{x} + b)$$

To think about: 2D contour plot.

High-Dimensional Polynomials

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

Under / Over fitting



Regularization

Discourage extreme fit

L_2 regularization

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

Cost function, which we minimize

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

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

Try $\lambda \in \{0, 10^{-3}, 10^{-2}, 0.1, 1, 10, 100\}$?

with $\underline{y}' = \begin{bmatrix} \underline{y} \\ \underline{0} \end{bmatrix} \begin{matrix} \updownarrow N \\ \updownarrow K \end{matrix}$, $\quad \Phi' = \overset{\xleftrightarrow{K}}{\begin{bmatrix} \Phi \\ \sqrt{\lambda}\, I_K \end{bmatrix}}$

$K \times 1$

K basis function

Identity matrix

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

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

means "fitted parameter"