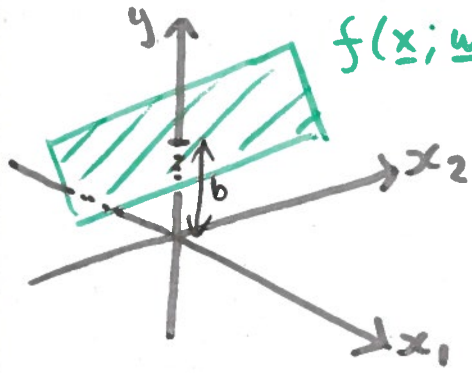


$$f(\underline{x};\underline{w},b) = \underline{w}^\top\underline{x} + b$$
$$= \underline{v}^\top\underline{\phi}(\underline{x})$$

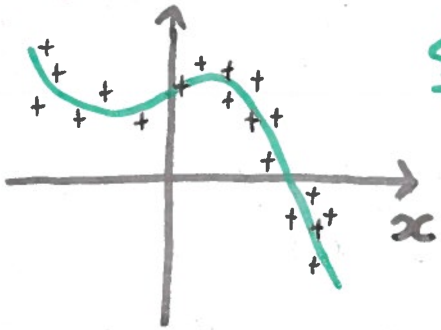
$$\underline{v} = \begin{bmatrix} b \\ \underline{w} \end{bmatrix}, \quad \underline{\phi}(\underline{x}) = \begin{bmatrix} 1 \\ \underline{x} \end{bmatrix}$$

$\underline{f} = \Phi\underline{v}$, choose $\underline{v}$ to minimize $(\underline{y}-\underline{f})^\top(\underline{y}-\underline{f})$

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

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

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

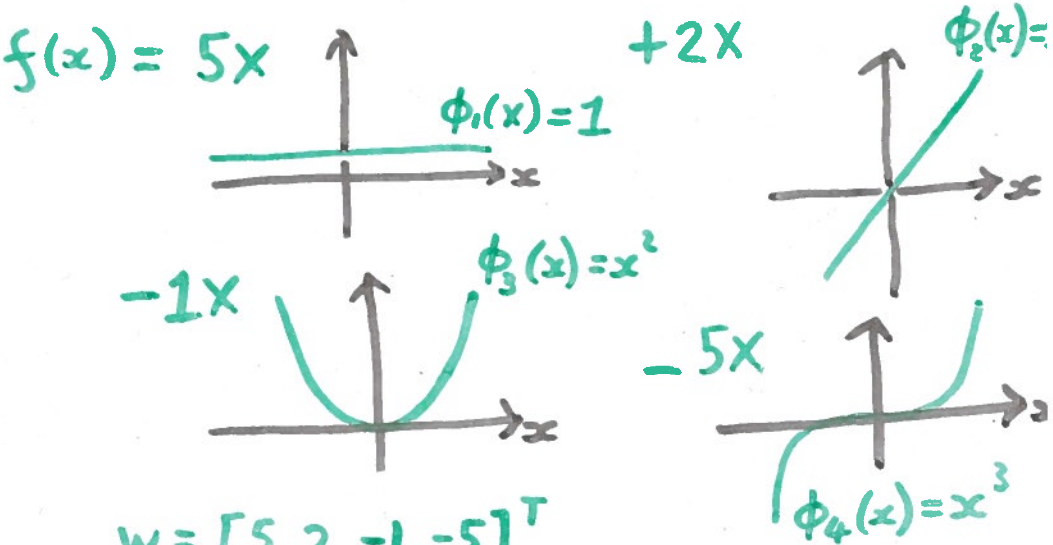


$\underline{\phi}(x) = [1 \;\; x \;\; x^2 \;\; x^3]^\top$

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

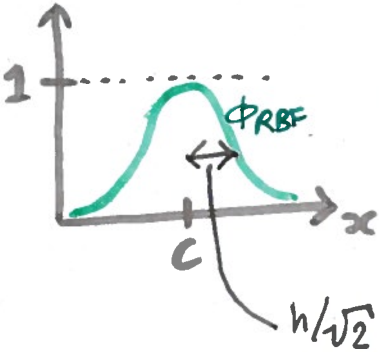


$\underline{w} = [5 \;\; 2 \;\; -1 \;\; -5]^T$

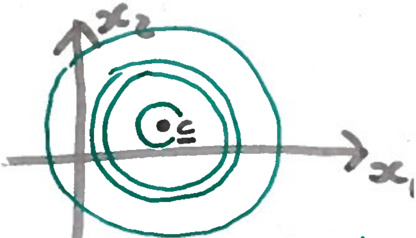
$$\underline{\phi}(\underline{x}) = [\phi_1(\underline{x}) \;\; \phi_2(\underline{x}) \;\; \cdots \;\; \phi_K(\underline{x})]^T$$

Basis function $\phi_k(\underline{x})$ could be anything...

Radial Basis Functions (RBFs)

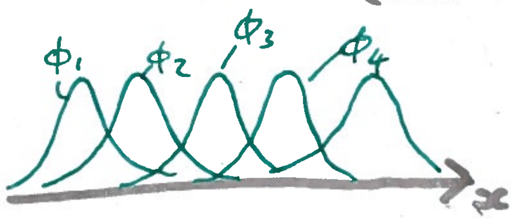


$$\phi_{RBF}(\underline{x}; \underline{c}, h) = \exp\left(-(\underline{x}-\underline{c})^\top(\underline{x}-\underline{c})/h^2\right)$$



Circular/Spherical contours of ϕ_{RBF}

Each ϕ_k uses different $\underline{c}$ (and/or h)



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

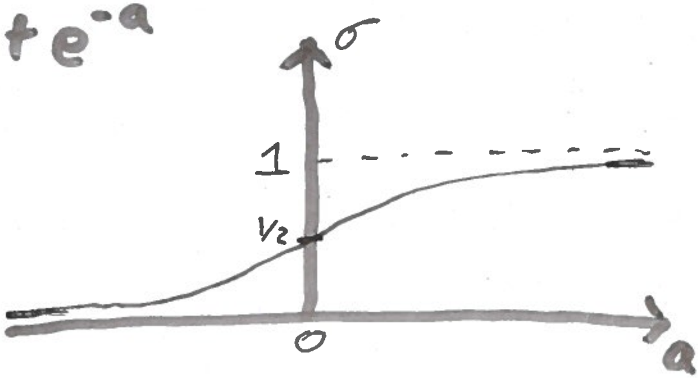
$f(x)$ with different weights

Logistic-Sigmoid function

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

↑

"activation"



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

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

[To think about: 2D contour plot]

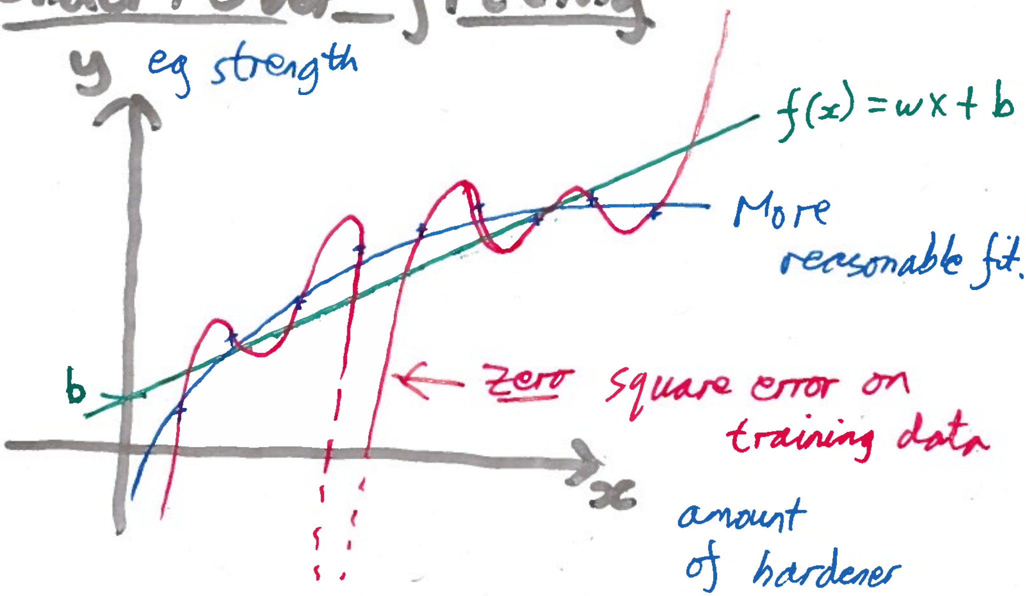
$$f = \underline{w}^T\underline{\phi}$$

Polynomials, multivariate input

$$\boldsymbol{\phi}(\mathbf{x}) = [\,1 \;\; x_1 \;\; x_2 \;\; x_3 \ldots$$
$$x_1^2 \;\; x_2^2 \;\; x_3^2 \ldots$$
$$x_1 x_2 \;\; x_2 x_3 \ldots$$
$$x_1^3 \;\; x_2^3 \;\; x_3^3$$
$$x_1^2 x_2 \;\; x_1^2 x_3 \;\; x_1 x_2 x_3$$
$$\ldots\,]$$

[Compare polynomial, RBF and σ basis functions]

Under/Over-fitting



Regularization

discourage extreme fits

L_2 regularization

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

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

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

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

If Φ $N \times K$

$\underline{y}'$ $(N+K) \times 1$

Φ' $(N+K) \times K$

[Can we fit λ by minimizing $E_\lambda(\underline{w})$?]