\[ f(x; w, b) = w^T x + b \]
\[ = v^T \phi(x) \]

\[ v = \begin{bmatrix} b \\ w \end{bmatrix} \]
\[ \phi(x) = \begin{bmatrix} 1 \\ x \end{bmatrix} \]

\[ f = \Xi v \]

Choose \( v \) to minimize
\[ (y - f)^T (y - f) \]

\[ f_n = f(x^{(m)}; v) \]

"\( v = \Xi \backslash y \)"

\[ \phi(x) = \begin{bmatrix} 1 & x & x^2 & x^3 \end{bmatrix}^T \]

Fit \( y \approx f = \Xi w \)

\[ f(x) = w_1 + w_2 x + w_3 x^2 + w_4 x^3 \]
Basis functions

\[ f(x) = 5x + 2x \]

\[ \phi_1(x) = 1 \]
\[ \phi_3(x) = x^2 \]
\[ \phi_4(x) = x^3 \]

\[ w = [5 \ 2 \ -1 \ -5]^T \]

\[ \phi(x) = [\phi_1(x) \ \phi_2(x) \ \ldots \ \phi_K(x)]^T \]

Basis function \( \phi_K(x) \) could be anything...
Radial Basis Functions (RBFs)

\[ \Phi_{RBF}(x; \xi, \eta) = \exp(-\frac{(x-\xi)^T(x-\xi)}{\eta^2}) \]

Each \( \Phi_k \) uses different \( \xi \) (and/or \( \eta \))

\[ f(x) = \sum_{k=1}^{5} w_k \Phi_k(x) \]

\( f(x) \) with different weights
Logistic Sigmoid function

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

"activation"

Basis function:

\[ \phi_\theta(x; \nu, b) = \sigma(\nu^T x + b) \]

[To think about: 2D contour plot]

\[ f = \nu^T \phi \]
Polynomials, multivariate input

\[ \phi(x) = [1 \ x_1 \ x_2 \ x_3 \ldots \ x_1^2 \ x_2^2 \ x_3^2 \ldots \ x_1 x_2 \ x_2 x_3 \ldots \ x_1^3 \ x_2^3 \ x_3^3 \ldots \ x_1^2 x_2 \ x_1^2 x_3 \ x_1 x_2 x_3 \ldots] \]

[Compare polynomial, RBF and \( \phi \) basis functions]
Under/Over-fitting

\[ y \]  eg strength

\[ f(x) = wx + b \]

More reasonable fit?

Zero square error on training data

Amount of hardener

Regularization

discourage extreme fits

L₂ regularization

\[ w^T w = \| w \|^2 \] should be small
\[ E_\lambda(w) = (y - \Phi w)^T(y - \Phi w) + \lambda w^T w \]

\[ = (y' - \Phi' w)^T(y' - \Phi' w) \]

with \( y' = \begin{bmatrix} y \\ 0 \end{bmatrix}, \Phi' = \begin{bmatrix} \Phi \\ \sqrt{k} I \end{bmatrix} \)

If \( \Phi \) \( N \times K \)
\( y' \) \( (N+k) \times 1 \)
\( \Phi' \) \( (N+k) \times K \)

[Can we fit \( \lambda \) by minimizing \( E_\lambda(w) \)?)]