UNIVERSITY OF AMSTERDAM
Informatics Institute

AMLAB
Amsterdam Machine Learning Lab

Machine Learning 1

Lecture 3.1 - Supervised Learning
Linear Regression With Basis Functions

Erik Bekkers

(Bishop 3.1)

Image credit: Kirillm | Getty Images

Slide credits: Patrick Forré and Rianne van den Berg

Three Statistical Learning Principles

Three general statistical learning principles to go from data to models (parametric predictive/proposal distributions):

I. Maximum likelihood

II. Maximum a posteriori

III. Bayesian prediction

$$p(t|x) = \mathcal{N}(t \,|\, y(x, \underline{w}), \beta^{-1})$$

Linear Regression

- Regression: $D = \{(\mathbf{x}_1, t_1), ..., (\mathbf{x}_N, t_N)\}$
  - Input variables $\underline{x}_i \in \mathbb{R}^D$
  - Target variables $t_i \in \mathbb{R}$

$$\underset{\sim}{\underline{w}} = \begin{pmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_D \end{pmatrix}, \quad \underset{\sim}{\underline{x}} = \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_D \end{pmatrix}$$

- Simplest linear model:

$$y(\mathbf{x}, \mathbf{w}) = w_0 + w_1 x_1 + w_2 x_2 + \dots + w_D x_D$$

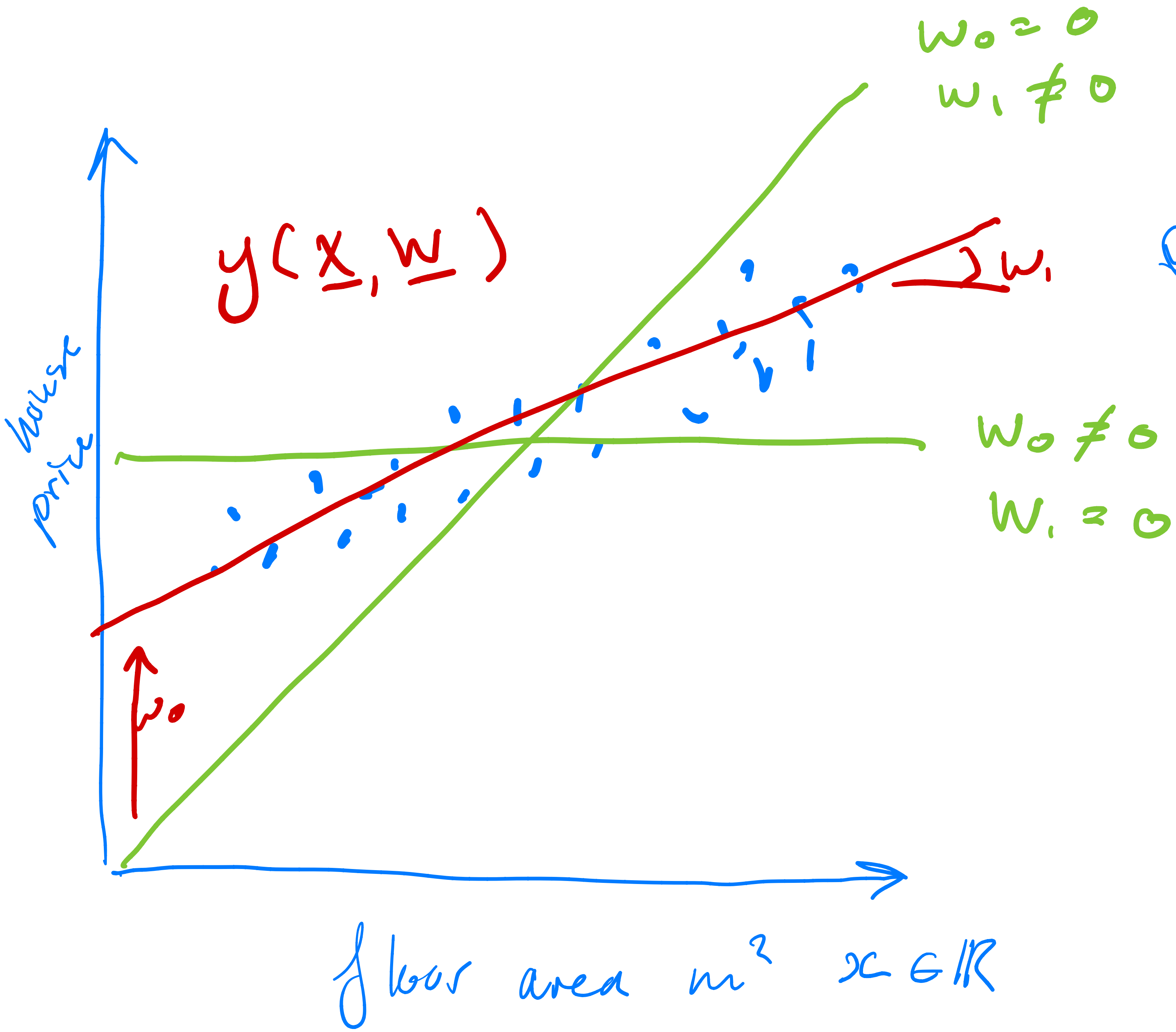
index refers to component within $\underline{x} \in \mathbb{R}^D$

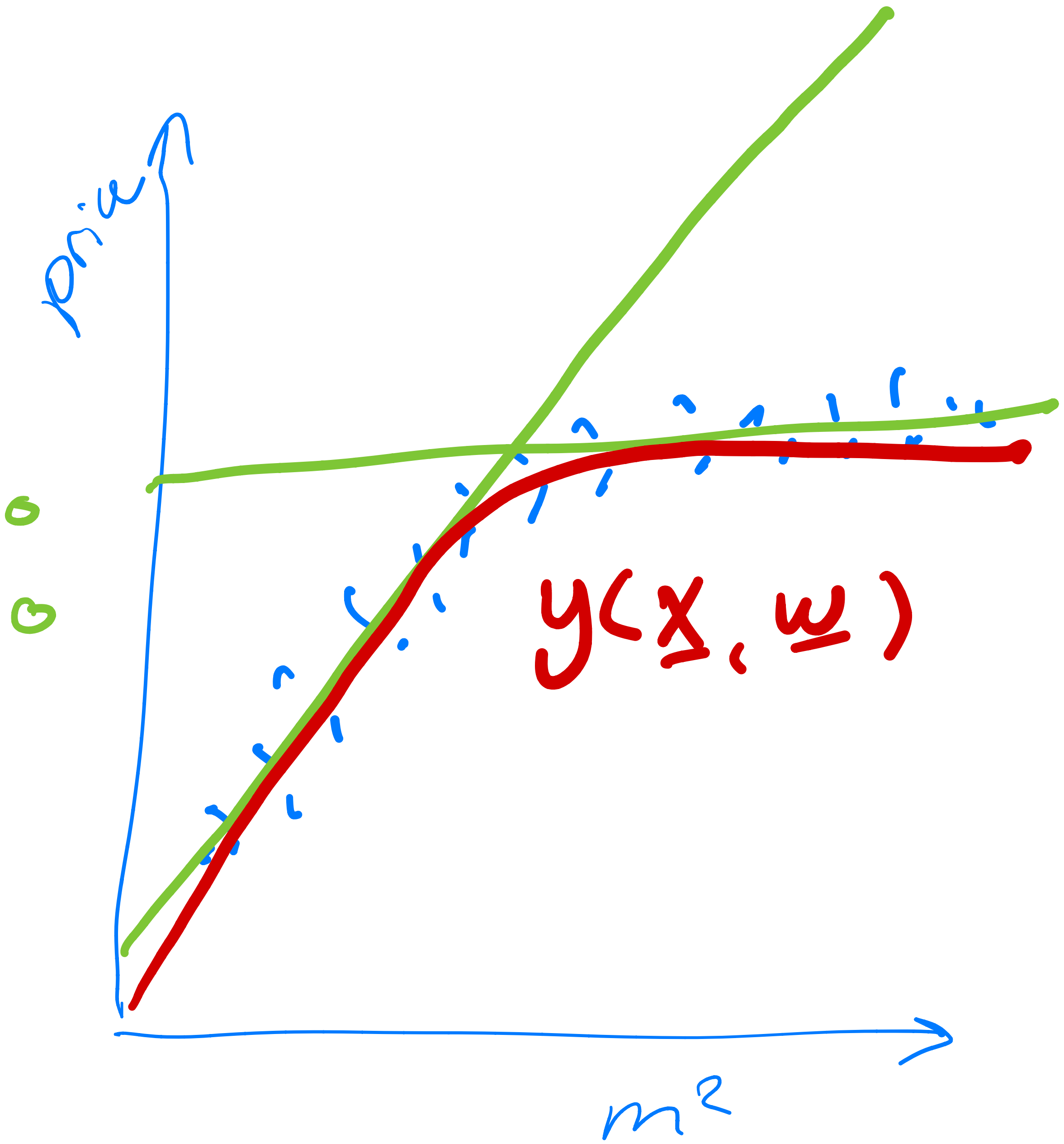
$$= w_0 + \underline{w}^T \underline{x}$$

$$= \tilde{\underline{w}}^T \tilde{\underline{x}}$$

Linear Regression

$$\underline{w} = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix}, \quad \underline{x} = \begin{pmatrix} 1 \\ x_1 \end{pmatrix}$$





Linear Basis Models

- Fix number of parameters *M* s.t. $\underline{w} \in \mathbb{R}^M$
- Choose *M - 1* basis functions/features of ***x***: $\phi_i(\underline{x}) \in \mathbb{R}$

  $\phi_i : \mathbb{R}^D \longrightarrow \mathbb{R}$ $\quad i = 1, \ldots, M-1$
- Approximation:

$$y(\mathbf{x}, \mathbf{w}) = w_0 + \sum_{i=1}^{M-1} w_i \, \phi_i(\underline{x})$$

- w_0 : bias

  $\underline{\phi} : \mathbb{R}^D \rightarrow \mathbb{R}^M$
- Set $\phi_0(\mathbf{x}) = 1$ such that $\boldsymbol{\phi}(\mathbf{x}) = [\phi_0(\underline{x}), \phi_1(\underline{x}), \ldots, \phi_{M-1}(\underline{x})]^T$

  $\hookrightarrow \phi_i : \mathbb{R}^D \rightarrow \mathbb{R}$

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

Example: Basis Functions (I) $(M = D)$

- Projection on input components : $\phi_i(\mathbf{x}) = x_i$

  for $\mathbf{x} = (x_1, x_2, ..., x_D)^T$: $y(\mathbf{x}, \mathbf{w}) = w_0 + \sum_{i=1}^{M} w_i \phi_i(x)$

  $= w_0 + \sum_{i=1}^{M} w_i x_i$

- *i*-power map for $x \in \mathbb{R}$: $\phi_i(x) = x^i$

$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 \cdot x^2 + w_3 x^3 \ldots$

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

$\underline{\phi}(x) = \begin{pmatrix} 1 \\ x \\ x^2 \\ x^3 \\ \vdots \end{pmatrix}$

Example: Basis Functions (III)

hyperparameters

- Gaussian basis functions: $\phi_i(\mathbf{x}) = \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T \Sigma_i^{-1} (\mathbf{x}-\boldsymbol{\mu}_i)\right)$

$\mathbf{x} \in \mathbb{R}^D$

$$y(\mathbf{x}, \mathbf{w}) = w_0 + \sum_{i=1}^{M-1} w_i \, e^{-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1} (x-\mu_i)}$$

hyperparameters

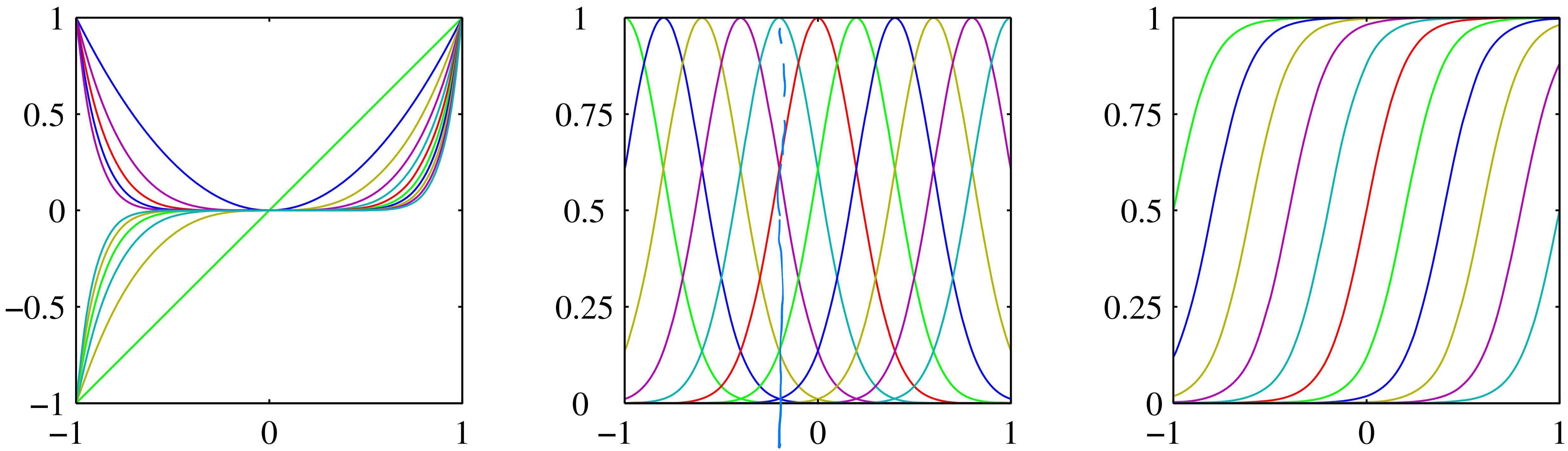
- Logistic sigmoid functions: $\phi_i(x) = \sigma\left(\frac{x-\mu_i}{s_i}\right)$ $\quad x \in \mathbb{R}$

with $\sigma(x) = \dfrac{1}{1+e^{-x}}$

$$y(x, \mathbf{w}) = w_0 + \sum_{i=1}^{M-1} w_i \, \sigma\left(\frac{x-\mu_i}{s_i}\right)$$

Example: Basis Functions (IV)

$\phi_i(x) = x^i$ $\phi_i(x) = e^{-\frac{1}{2\sigma^2}(x-\mu_i)^2}$ $\phi_i(x) = \sigma\left(\frac{x-\mu_i}{s}\right)$



μ_i

$x \in \mathbb{R}$

Figure: Example of basis functions (Bishop 3.1)