Regression

- Regression: prediction. line fitting $y = mx + c$
- General: t: 1-D target variable, what is observed
$$\boxed{t = y(\mathbf{x}) + \varepsilon}$$
$\mathbf{x}$ is the $M-1$-dimensional input data
$y(\cdot)$: 1-D function of $(M-1)-$dim input: the model
ε: noise ($\sim$ can't model, sometimes modelled)
- Reconciliation: may not be able to model all well
- Simple 2-D case: y: an implicit function of $\mathbf{x}$.
e.g., $f(x,y) = ax + by + c = 0$, or $w_2x_2 + w_1x_1 + w_0 = 0$
- $y(\mathbf{x}) = y(\mathbf{x}, \mathbf{w}) = w_2x_2 + w_1x_1 + w_0x_0$. $x_0 = 1$, w_0: bias
- Written equivalently in two ways:
$y(\mathbf{x}, \mathbf{w}) = \mathbf{w}^T\mathbf{x}$, for $(D+1) = M-$ dim data, $x_0 = 1$, or
$y(\mathbf{x}, \mathbf{w}) = \mathbf{w}^T\mathbf{x} + w_0$, for $D = (M-1)-$ dim data

Home Page

Title Page

Contents

Page 4 of 30

Go Back

Full Screen

Close

Quit

- $y(\mathbf{x},\mathbf{w}) = w_0 x_0 + \ldots w_{M-1} x_{M-1} = \sum_{j=0}^{M-1} w_j x_j$
- Generalising to scalar basis functions $\phi_j(\mathbf{x})$:
- $y(\mathbf{x},\mathbf{w}) = \sum_{j=0}^{M-1} w_j \phi_j(\mathbf{x}) = w_o \phi_0(\mathbf{x}) + \ldots w_{M-1} \phi_{M-1}(\mathbf{x})$
- Model: linear combo of fixed basis fns (lin/non-lin)

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

- Not most general, but practically imp! Examples:
  - Polynomial basis fns: x^j: global, unlike splines
  - Gaussian basis fns
  - Sigmoidal basis fns
  - Fourier basis fns
  - Wavelet basis fns localised in space & freq

Home Page

Title Page

Contents

◀◀ ▶▶

◀ ▶

Page 5 of 30

Go Back

Full Screen

Close

Quit

Maximum Likelihood, Least Sq

- $\boxed{t = y(\mathbf{x}) + \varepsilon}$; $\boxed{t = y(\mathbf{x}, \mathbf{w}) + \varepsilon}$
- t: target variable $y(\cdot)$: deterministic fn (model)
- $\mathbf{x}$: input, $\mathbf{w}$: parameters, ε: noise
- ε: take as the unmodelled part: the residue, or
- ... model ε as well. Common: $\varepsilon = \mathcal{N}(0, \sigma^2)$
- $t = \mathcal{N}(y(\cdot), \sigma^2)$, Mean: $y(\mathbf{x}, \mathbf{w})$, variance: σ^2
- If no $y(\cdot)$, t usually 0, or small +/-: weighing m/c noise: zero error, offset: $y(\cdot)$
- $p(t|\mathbf{x}, \mathbf{w}, \sigma^2) = p(t|\mathbf{w}, \sigma^2) = \mathcal{N}(t|y(\mathbf{x}, \mathbf{w}), \sigma^2)$

 Copyright ©Sumantra Dutta Roy