Seoul National University
Dept. of Civil and Environmental Engineering

Instructor: Junho Song
junhosong@snu.ac.kr

**457.646 Topics in Structural Reliability**

**In-Class Material: Class 27**

## ◎ Basic formulation of RS models

Two approaches: regression ⇒ use assumed mathematical model & fit it to data

e.g. $\eta(\mathbf{x}) = \sum_{i=1}^{p} \theta_i x_i^m$

Interpolation ⇒ Interpolate using nearby data points

e.g. K-nearest points

<u>Regression</u>

True response of $g(\mathbf{x})$: $Z(\mathbf{x})$

$$Z(\mathbf{x}) = \eta(\underbrace{\theta_1, \cdots, \theta_p}_{\text{Model parameters}}; \mathbf{x}) + \varepsilon$$

- $\theta_1, \cdots, \theta_p$: Model parameters
- $\mathbf{x}$: Input
- $\varepsilon$: <u>Zero mean</u> (random) error term ⇒ $E[z-\eta] = E[\varepsilon] = 0$

"unbiased" model

How to find $\boldsymbol{\theta}$? What do data tell us?

Ref: Tipping, M.E. (2004)

"Bayesian inference: an introduction to principles and practice in machine learning" Advanced lectures on machine learning, pp.41-62

(Free codes and papers at miketipping.com)

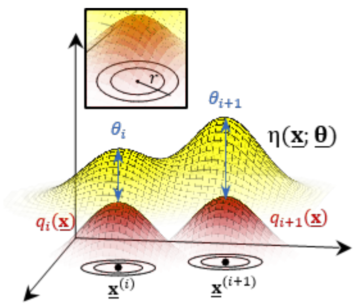
$$\eta = \theta_1 \exp(x) + \theta_2 \ln x + \theta_3 \cdots$$

## ◎ <u>Linear</u> models (Linear in )

Find $Z = \eta(\mathbf{x};\boldsymbol{\theta}) + \varepsilon$

$$= \sum_{i=1}^{p} \theta_i \, q_i(\mathbf{x}) + \varepsilon$$

- $\theta_i$: Model Parameter
- $q_i(\mathbf{x})$: Basis Function (Shape function)




e.g. $q_i(\mathbf{x}) \propto$ PDF of $N(\mathbf{x}^{(i)}, r^2\mathbf{I})$

from $\{\mathbf{x}^{(i)}, Z^{(i)}\}$, $i = 1, \cdots, m$

$\mathbf{Z} = \mathbf{Q}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$

!!!

$$\begin{Bmatrix} Z^{(1)} \\ Z^{(2)} \\ \vdots \\ Z^{(m)} \end{Bmatrix} = \begin{bmatrix} q_1(\mathbf{x}^{(1)}) & \cdots & \cdots & q_p(\mathbf{x}^{(1)}) \\ \vdots & & & \\ q_1(\mathbf{x}^{(m)}) & \cdots & \cdots & q_p(\mathbf{x}^{(m)}) \end{bmatrix} \begin{Bmatrix} \theta_1 \\ \vdots \\ \theta_p \end{Bmatrix} + \begin{Bmatrix} \varepsilon^{(1)} \\ \vdots \\ \varepsilon^{(m)} \end{Bmatrix}$$

$m \times 1$ $\quad$ $m \times p$ $\quad$ $p \times 1$ $\quad$ $m \times 1$

Five approaches (Tipping 2004)

① "Least-Square" Approximation (classic)

⇒ Minimize sum of squared errors

$$E_D = \frac{1}{2}\sum_{i=1}^{m}(Z^{(i)} - \eta(\mathbf{x}^{(i)}, \theta))^2$$

$$= \frac{1}{2}(\mathbf{Z} - \mathbf{Q}\boldsymbol{\theta})^T(\mathbf{Z} - \mathbf{Q}\boldsymbol{\theta})$$

$$= \frac{1}{2}\mathbf{Z}\mathbf{Z}^T + \frac{1}{2}(\mathbf{Q}\boldsymbol{\theta})^T(\mathbf{Q}\boldsymbol{\theta}) - \mathbf{Z}^T\mathbf{Q}\boldsymbol{\theta}$$

$$\frac{\partial E_D(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\mathbf{Z}^T\mathbf{Q} + (\mathbf{Q}\boldsymbol{\theta})^T\mathbf{Q} = 0$$

Solve for $\boldsymbol{\theta}$,

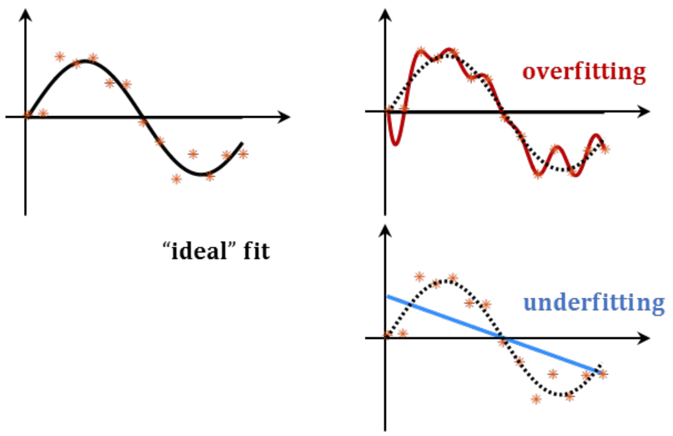
$$\boldsymbol{\theta}_{LS} = (\mathbf{Q}^T\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{Z}$$

※ over-fitting?

e.g. $Z = \sin x + \varepsilon$

$\sin x \rightarrow$ true model, $\varepsilon \rightarrow$ noise

Figure 1 in Tipping (2004)



Instructor: Junho Song
junhosong@snu.ac.kr

② Regularization (by giving penalty on large $\boldsymbol{\theta}$)

$$\hat{E}(\boldsymbol{\theta}) = E_D(\boldsymbol{\theta}) + \lambda \; E_W(\boldsymbol{\theta})$$

Standard choice

$$E_W(\boldsymbol{\theta}) = \frac{1}{2}\sum_{i=1}^{p} \theta_i^2$$

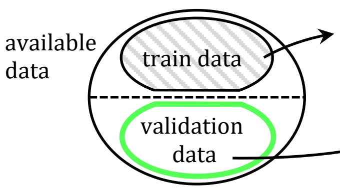
regularization parameter $\lambda \uparrow$ Discourage large value of $\boldsymbol{\theta}$

$\Rightarrow$ Smooth function

$$\frac{\partial E_D(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0 \;\Rightarrow\; \boldsymbol{\theta}_{PLS} = (\mathbf{Q}^T\mathbf{Q} + \lambda \mathbf{I})^{-1}\mathbf{Q}^T\mathbf{Z}$$

※ Appropriate value of $\lambda$?

A common approach: Use "validation" data




Find $\boldsymbol{\theta}_{PLS}$ for a given $\lambda$, $\boldsymbol{\theta}_{PLS}(\lambda)$

Construct surrogate $\eta(\mathbf{x};\lambda) = \sum_{i=1}^{p} \theta_i^{PLS}(\lambda) q_i(\mathbf{x})$

Compute error against validation data

$\varepsilon = Z - \eta$ at each $\mathbf{x}^{(i)} \in$ validation data set

and choose $\lambda$ that makes $\frac{1}{2}\Sigma(Z-\eta)^2$ minimum

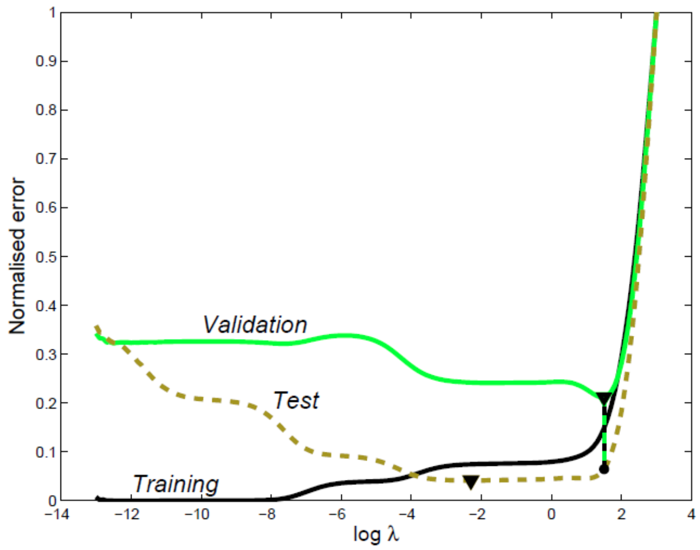



**Fig. 3.** Plots of error computed on the separate 15-example training and validation sets, along with 'test' error measured on a third noise-free set. The minimum test and validation errors are marked with a triangle, and the intersection of the best $\lambda$ computed via validation is shown.

Instructor: Junho Song
junhosong@snu.ac.kr

※ Probabilistic Regression

$Z = \eta + \varepsilon$ !

e.g. $\varepsilon \sim N(0, \sigma^2)$ $\quad \therefore Z \sim N(\eta, \sigma^2)$

Using this information one can construct likelihood function

$$L(\mathbf{Z} | \mathbf{x}, \boldsymbol{\theta}, \sigma^2) = \prod_{i=1}^{n} f(Z^{(i)} | \mathbf{x}^{(i)}, \boldsymbol{\theta}, \sigma^2)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[ -\frac{\{Z^{(i)} - \eta(\mathbf{x}^{(i)}; \boldsymbol{\theta})\}^2}{2\sigma^2} \right]$$

③ Maximum Likelihood Estimation

Find $\boldsymbol{\theta}$ that maximizes L( ) ⇔ Find $\boldsymbol{\theta}$ that minimizes –lnL( )

$$-\ln L(\quad) = \frac{n}{2}\ln(2\pi\sigma^2) + \frac{1}{2\sigma^2}\sum_{i=1}^{n}\{Z^{(i)} - \eta(\mathbf{x}^{(i)}, \boldsymbol{\theta})\}^2$$

→ $E_D(\boldsymbol{\theta})$
⇒ error measure for $\boldsymbol{\theta}_{LS}$

Therefore, MLE based on s.i. ________ error assumption (i.e. $\varepsilon \sim N(\quad)$)

Gives $\quad \boldsymbol{\theta}_{MLE} = \boldsymbol{\theta}_{LS}$

(cf. Assuming errors are dependent? $\varepsilon \sim N(\mathbf{0}, \boldsymbol{\Sigma})$)

$$\rho_{ij} = \exp\left( -\frac{\left\| \mathbf{x}^{(i)} - \mathbf{x}^{(j)} \right\|}{L} \right) \Rightarrow$$ "Kriging" Method (Satner et al. 2003)

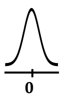
※ Bayesian Methods $f = c \cdot L \cdot p$

Introduce a prior distribution

$$p(\boldsymbol{\theta} | \alpha) = \prod_{i=1}^{p} \left( \frac{\alpha}{2\pi} \right)^{1/2} \exp\left\{ -\frac{\alpha}{2}\theta_i^2 \right\}$$

$$= \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi}\ {}^{1}\!/\!\sqrt{\alpha}} \exp\left\{ -\frac{\theta_i^2}{2({}^{1}\!/\!_{\alpha})} \right\}$$

(degree of belief about smooth model)

$\alpha \uparrow$ Variability reduces  ⇒ certain that $\theta$ is around 0

⇒Become smooth

$\therefore \ \alpha \propto \lambda$

Instructor: Junho Song
junhosong@snu.ac.kr

④ Maximum a posteriori (MAP) estimation (a Bayesian "shortcut")

$$f = c \cdot L \cdot p$$

$$P(\boldsymbol{\theta}|\mathbf{Z},\alpha,\sigma^2) = c \cdot L(\mathbf{Z}|\boldsymbol{\theta},\sigma^2) \cdot p(\boldsymbol{\theta}|\alpha)$$

Posterior Likelihood function prior

Find $\boldsymbol{\theta}$ where $P(\boldsymbol{\theta}|\mathbf{Z},\alpha,\sigma^2)$ is maximum

e.g. Normal s.i errors $\varepsilon$, $Z \sim N(\eta,\sigma^2)$

$$-\ln(f) = \frac{1}{2\sigma^2}\sum_{i=1}^{n}\{Z^{(i)} - \eta(\mathbf{x}^{(i)};\boldsymbol{\theta})\}^2 + \frac{\alpha}{2}\sum_{i=1}^{p}\theta_i^2$$

$\lambda = \alpha\,\sigma^2$

$$-\sigma^2\ln(f) = \frac{1}{2}\sum_{i=1}^{n}\{Z^{(i)} - \eta(\mathbf{x}^{(i)};\boldsymbol{\theta})\}^2 + \frac{\alpha\sigma^2}{2}\sum_{i=1}^{p}\theta_i^2$$

$E_D(\boldsymbol{\theta})$ $E_W(\boldsymbol{\theta})$ $\frac{1}{2}\lambda$

the same as

※ $\alpha$, $\sigma^2$ ? no need to bother w/ Bayesian?

⑤ **Full** Bayesian ("Marginalization")

integrate $P(\mathbf{Z}|\boldsymbol{\theta},\alpha,\sigma^2)$ over all $\boldsymbol{\theta}$

$$P(\mathbf{Z}) = \int P(\mathbf{Z}|\boldsymbol{\theta})\cdot P(\boldsymbol{\theta})d\boldsymbol{\theta}$$

Focus on

Total probability theorem

$$P(\mathbf{Z}|\alpha,\sigma^2) = \int P(\mathbf{Z}|\boldsymbol{\theta},\alpha,\sigma^2)\cdot P(\boldsymbol{\theta}|\alpha,\sigma^2)d\boldsymbol{\theta}$$

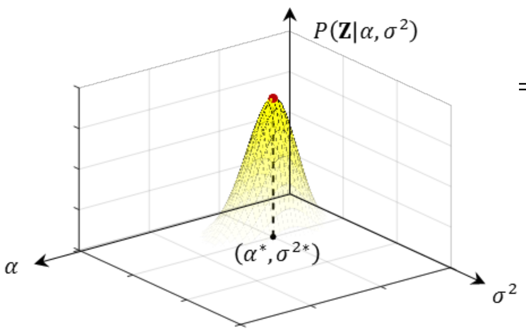
Simplified to

$$= \int P(\mathbf{Z}|\boldsymbol{\theta},\sigma^2)\cdot P(\boldsymbol{\theta}|\alpha)d\boldsymbol{\theta}$$

→ Closed-form available:

$f_N(\mathbf{Z},\alpha,\sigma^2)$ (Eq. 23 in Tipping, 2004)

※ $P(\mathbf{Z}|\alpha,\sigma^2)$ : Probability that you will observe $\mathbf{Z}$ for given $\alpha,\sigma^2$




⇒ Find $\alpha$ & $\sigma^2$ that maximizes $P(\mathbf{Z}|\alpha,\sigma^2)$

(i.e. Let data $\mathbf{Z}$ tell us the optimal point $\alpha^*,\sigma^{2*}$)

Instructor: Junho Song
junhosong@snu.ac.kr

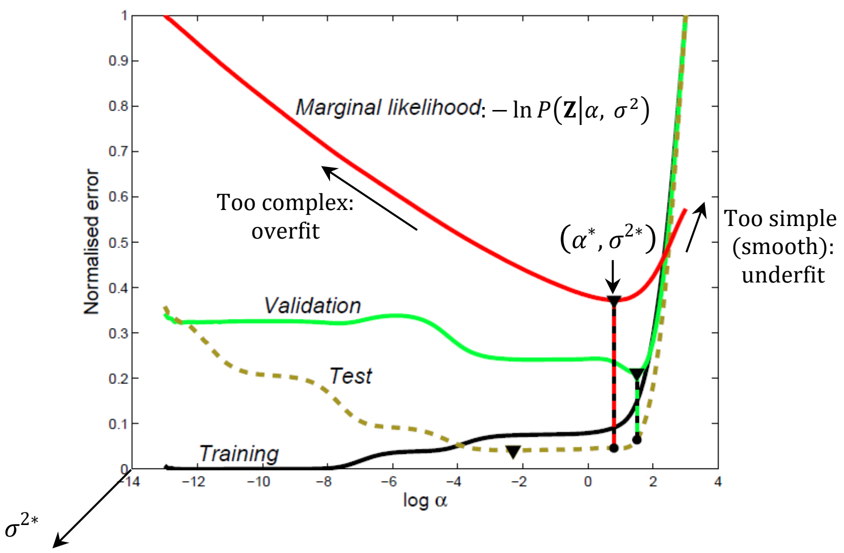



**Fig. 5.** Plots of the training, validation and test errors of the model as shown in Figure 3 (with the horizontal scale adjusted appropriately to convert from $\lambda$ to $\alpha$) along with the negative log marginal likelihood evaluated *on the training data alone* for that same model. The values of $\alpha$ and test error achieved by the model with highest marginal likelihood (smallest negative log) are indicated.

☆ **Okham's Razar** (or the law of parsimony)**:**

**"model should be no more complex than is sufficient to explain the data"**

CRC CH.19 RS
→DOE
→$q_i(\mathbf{x})$

## ◎ Other RS or UQ methods

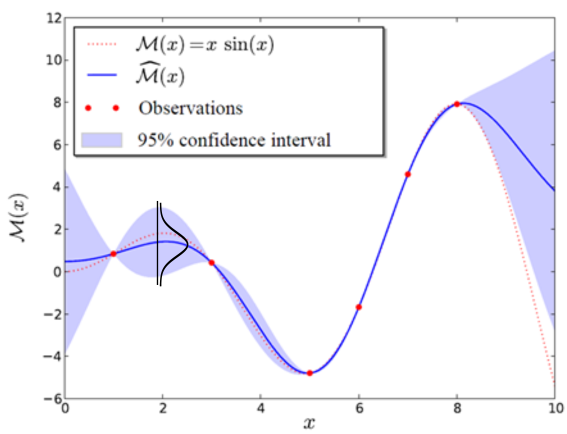
① Kriging (Santner et al. 2003)

(Dubourg et al. 2010 IFIP)

$\varepsilon \sim N(\mathbf{0}, \mathbf{\Sigma})$

e.g. $\rho_{ij} = \exp\left(-\frac{\left\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\right\|}{L}\right)$

- coincides at each point
- Interpolate b/w each point
- Can quantify confidence
- Regularization




(Dubourg et al. 2011)

Instructor: Junho Song
junhosong@snu.ac.kr

② Dimension Reduction (Rahman & Xu, 2004; Xu & Rahman 2004)

$$g(\mathbf{x}) \rightarrow g(\hat{\mathbf{x}}) = \sum_{i=1}^{n} g(\mu_1, \cdots, \mu_{i-1}, x_i, \mu_{i+1}, \cdots, \mu_n) - (n-1) g(\mu_1, \cdots, \mu_n)$$

$\Downarrow$

$$E[(g(x))^m] \cong E[(\hat{g}(x))^m] = \int (\hat{g}(x))^m \underline{f_{\mathbf{x}}(\mathbf{x})} d\mathbf{x} \quad \nearrow \Pi \varphi(x_i)$$

Transform to s.i. space; Multivariate Integral $\Rightarrow$ Multiple univariate Integral

③ Polynomials chaos (a good review by Eldred et al. 2008)

$$R = a_0 B_0 + \sum_{i_1=1}^{\infty} a_{i1} B_1(\zeta_{i1}) + \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{\infty} a_{i1,i2} B_2(\zeta_{i1} \zeta_{i2}) + \cdots$$

$$= \sum_{j=0}^{p} \alpha_j \psi_j(\zeta)$$ → Orthogonal bases for given types of r.v's distribution

$$\alpha_j = \frac{< R, \psi_j >}{< \psi_j^2 >} = \frac{\int R \psi_j f(\zeta) d\zeta}{< \psi_j^2 >}$$

(numerator) → Important sampling, etc.

(denominator) → closed form available