ASE-5010/5016	Advanced methods for data-driven modeling and analysis	Hannu Koivisto
TUT/ASE	Exam	11.12.2013

1. Describe shortly next terms from the regression point of view
   (a) generalization capability
   (b) regularization
   (c) bias-variance decomposition

2. List three common clustering methods and give shortly their relative strengths

3. Are there cases when MLP networks are clearly a better alternative for function approximation than ANFIS? What about the opposite?

4. Compare Mamdani and Takagi-Sugeno type of fuzzy reasoning. What are the basic differences and what are the benefits gained when applying either?

5. We have N measurements $\{x_n, y_n\}$. We would like to fit a linear-in-the-parameters basis function model $y_n = \boldsymbol{w}^T \boldsymbol{\phi}(x_n)$, where the vector $\boldsymbol{w}$ contains the parameters. The measurements set is already re-organized using the design matrix notation i.e. $\boldsymbol{Y} = \boldsymbol{y}_{1:N} = \boldsymbol{\Phi}\boldsymbol{w}$.

   Use the Bayesian approach to obtain parameter estimates for $\boldsymbol{w}$ i.e. compute the corresponding posterior distribution $p(\boldsymbol{w}|\boldsymbol{y}_{1:N}, \boldsymbol{x}_{1:N}, \beta)$. The measurement is assumed Gaussian with known variance as $\sigma^2 = 1/\beta$. Assume parameter prior distribution as

$$p(\boldsymbol{w}|\boldsymbol{\alpha}) = \mathcal{N}(\boldsymbol{w}|\boldsymbol{0}, \alpha^{-1}\boldsymbol{I})$$

   You could apply the transformation equations below.

Linear transformation of multivariate Gaussian

Given two probability densities

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1}) \quad (1)$$

$$p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{L}^{-1}) \quad (2)$$

we have

$$p(\mathbf{y}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{L}^{-1} + \mathbf{A}\boldsymbol{\Lambda}^{-1}\mathbf{A}^T) \quad (3)$$

$$p(\mathbf{x}|\mathbf{y}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\Sigma}\{\mathbf{A}^T\mathbf{L}(\mathbf{y} - \mathbf{b}) + \boldsymbol{\Lambda}\boldsymbol{\mu}\}, \boldsymbol{\Sigma}) \quad (4)$$

where

$$\boldsymbol{\Sigma} = (\boldsymbol{\Lambda} + \mathbf{A}^T\mathbf{L}\mathbf{A})^{-1}$$