National Chiao Tung University, Department of Computer Science
IOC5184 Deep Learning and Practice–Final Exam

Instructors: Dr. Wen-Hsiao Peng, Dr. I-Chen Wu, and Dr. Yong-Sheng Chen
Date: Thursday, August 31, 2017
Time: 1:20pm – 4:30pm
Format: Open book
Instructions:
1) You may give your answers in Chinese or English.
2) Please give your answers in succinct phrases or point form.
3) Please write your answers clearly (with explicit denotation of labels and symbols used).

1. (50 pts) In the **linear regression** problem with **Bayesian** statistics, the settings are as follows:

$$\text{Visible variables: } y_i = \phi(x_i)^T \boldsymbol{w} + \varepsilon_i, i = 1, 2, \ldots, N$$
$$\text{Latent variables: } \boldsymbol{w} = (w_1, w_2, \ldots, w_M)$$

where ε_i are independently and identically distributed Gaussian noises, and independent of $\boldsymbol{w}$ with[1]

$$p(\varepsilon_i) = \mathcal{N}(\varepsilon_i; 0, \beta^{-1}), i = 1, 2, \ldots, N$$
$$p(\boldsymbol{w}) = \mathcal{N}(\boldsymbol{w}; \boldsymbol{\mu}, \lambda^{-1}\boldsymbol{I})$$

(a) (4 pts) Draw a directed graphical model to capture the independence between the visible variables $y_i, i = 1, 2, \ldots, N$ and the latent variables $w_j, j = 1, 2, \ldots, M$.

(b) (2 pts) Show that $p(\boldsymbol{y}), \boldsymbol{y} = (y_1, y_2, \ldots, y_N)$ is generally NOT factorial, using the d-separation rules.

(c) (2 pts) Show that $p(\boldsymbol{w}|\boldsymbol{y}), \boldsymbol{y} = (y_1, y_2, \ldots, y_N)$ is generally NOT factorial, using the d-separation rules.

(d) (7 pts) Show that the posterior $p(\boldsymbol{w}|\boldsymbol{y}), \boldsymbol{y} = (y_1, y_2, ..., y_N)$ is given by

$$p(\boldsymbol{w}|\boldsymbol{y}) = \mathcal{N}(\boldsymbol{w}; \boldsymbol{u}_N, \boldsymbol{\Lambda}_N^{-1}),$$

with

$$\boldsymbol{\Lambda}_N = \lambda \boldsymbol{I} + \beta \boldsymbol{\Phi}^T \boldsymbol{\Phi}$$
$$\boldsymbol{u}_N = \boldsymbol{\Lambda}_N^{-1}(\lambda \boldsymbol{\mu} + \beta \boldsymbol{\Phi}^T \boldsymbol{y})$$
$$\boldsymbol{\Phi} = \begin{bmatrix} \phi(x_1)^T \\ \phi(x_2)^T \\ \vdots \\ \phi(x_N)^T \end{bmatrix}$$

(e) Assuming that $\boldsymbol{\mu}, \lambda, \beta$ are unknown model parameters, use the EM learning algorithm to derive the iterative update equations for these parameters.

---

[1] n-dimensional Gaussian: $p(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1}) \triangleq \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Lambda}^{-1}|^{\frac{1}{2}}} \exp(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Lambda}(\boldsymbol{x} - \boldsymbol{\mu}))$