PRML chapter14

1. Bayesian Model Averaging
   - The marginal distribution of the data is

$$p(\mathbf{X}) = \sum_{h=1}^{H} p(\mathbf{X}|h)p(h)$$

     - p(h) is a probability of the model is chosen
     - p(X|h) is a probability X is generated given h
     - In contrast, in joint models, z means which model causes the data.
2. Committees
   - average the approximation from the each models
   - Though the data set is unique, we use bootstrap and obtain some data sets.
   - $y_{\mathrm{COM}}(\mathrm{x}) = \frac{1}{M}\sum_{m=1}^{M} y_m(\mathrm{x})$
3. Boosting
   - AdaBoost
     - initialize data coefficient w by uniformaly
     - loop the following calculation, minimizing the weighted error function

$$J_m = \sum_{n=1}^{N} w_n^{(m)} I\left(y_m\left(\mathbf{x}_n\right) \neq t_n\right)$$

       and evaluate the quantity

$$\epsilon_m = \frac{\sum_{n=1}^{N} w_n^{(m)} I\left(y_m\left(\mathbf{x}_n\right) \neq t_n\right)}{\sum_{n=1}^{N} w_n^{(m)}}$$

$$\alpha_m = \ln\left\{\frac{1-\epsilon_m}{\epsilon_m}\right\}$$

       then, update the weighted coefficient

$$w_n^{(m+1)} = w_n^{(m)} \exp\left\{\alpha_m I\left(y_m\left(\mathbf{x}_n\right) \neq t_n\right)\right\}$$

     - make predictions

$$Y_M(\mathbf{x}) = sign\left(\sum_{m=1}^{M} \alpha_m y_m(\mathbf{x})\right)$$

   (a) Minimizing exponential error

- exponential error is defined as

$$E = \sum_{n=1}^{N} \exp\left\{-t_n f_m\left(\mathbf{x}_n\right)\right\}$$

under

$$f_{\mathrm{m}}(\mathrm{x}) = \frac{1}{2}\sum_{l=1}^{m} \alpha_l y_l(\mathrm{x})$$

- we want to minimize the exponential error. First minimize with respect to α_m and y_m under $\alpha_1 ... \alpha_{m-1}$ and $y_1 ... y_{m-1}$ are fixed
- we obtain the above formulation in AdaBoost

(b) Error functions for boosting

- define error function as

$$E_{\mathbf{x},t}[\exp\{-ty(\mathbf{x})\}] = \sum_t \int \exp\{-ty(\mathbf{x})\} p(t|\mathbf{x}) p(\mathbf{x}) \mathrm{d}\mathbf{x}$$

- The difference between cross-entropy error and exponential error

4. Tree-based Models
   - partition the input space into cuboid region
5. Conditional Mixture Models

(a) Mixtures of linear regression models

- log loglikelihood function is written as

$$p(t|\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(t|\mathbf{w}_k^{\mathrm{T}}\boldsymbol{\phi}, \beta^{-1})$$

and we could appy EM algorithms

(b) Mixtures of logistic models

(c) likelihood function is defined as

$$p(\mathbf{t}|\boldsymbol{\theta}) = \prod_{n=1}^{N} \left(\sum_{k=1}^{K} \pi_k y_{nk}^{t_n} \left[1 - y_{nk}\right]^{1-t_n} \right)$$

and we could apply EM algorithms

(d) Mixtures of experts

- We can further increase the capability of such models by allowing the mixing coefficients themselves to be functions of the input variable