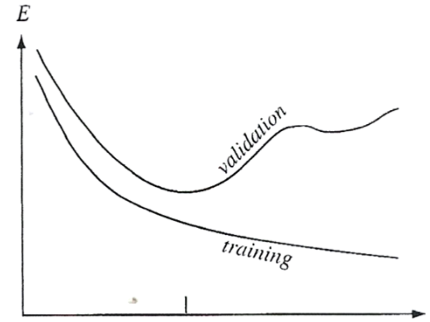
Programming & Experiment Assignment

- Classification experiment using GMMs
  - Implement an EM algorithm that can learn the parameters of GMMs.
  - Train GMMs for binary classification using the training data.
    - Use the K-fold cross validation method to find an optimal number of mixture components.

	$\mathcal{D} = \{x^t, r^t\}_{t=1}^N$				Training Data	Validation Data
1	$\mathcal{D}_1$	$\mathcal{D}_2$	$\cdots$	$\mathcal{D}_K$	$\mathcal{D}_2 \cup \mathcal{D}_3 \cup \mathcal{D}_K$	$\mathcal{D}_1$
2	$\mathcal{D}_1$	$\mathcal{D}_2$	$\cdots$	$\mathcal{D}_K$	$\mathcal{D}_1 \cup \mathcal{D}_3 \cup \mathcal{D}_K$	$\mathcal{D}_2$
$\vdots$			$\cdots$		$\vdots$	$\vdots$
K	$\mathcal{D}_1$	$\mathcal{D}_2$	$\cdots$	$\mathcal{D}_K$	$\mathcal{D}_1 \cup \mathcal{D}_2 \cup \mathcal{D}_{K-1}$	$\mathcal{D}_K$



  - Report the error rate on the test data when using the optimal number of mixture components.

- Submit the followings.
  1. Source code and documentation (40 points)
  2. Description of the training, validation, and testing procedures (10 points)
  3. Explanation on the initialization method and termination condition (10 points)
  4. Error curve graph for the K-fold cross validation experiments (30 points)
  5. The error rate on the test data (10 points)

Data File Format

- Training/test data file format
  - $$\begin{array}{ccccc} x_1^1 & x_2^1 & \cdots & x_d^1 & r^1 \\ x_1^2 & x_2^2 & \cdots & x_d^2 & r^2 \\ \vdots & & & & \\ x_1^N & x_2^N & \cdots & x_d^N & r^N \end{array}$$

- Sample file

```
-2.120279e+001 -5.292185e+000 ... -1.098583e+000 0
-2.267072e+001 -6.845204e+000 ... -3.950064e+000 0
-2.177675e+001 -7.340329e+000 ... -3.973404e+000 0
 ⋮
-2.035374e+001 -5.470649e+000 ... -3.804142e+000 1
-2.103584e+001 -5.863925e+000 ... -3.812235e+000 1
-2.162024e+001 -4.248268e+000 ... -3.703966e+000 1
 ⋮
```

Implementation Issues

- Finite precision computation of $P\left(g_k \middle| \boldsymbol{x}^t, \theta^i\right)$
  - $$P\left(g_k \middle| \boldsymbol{x}^t, \theta^i\right) = \frac{P(\boldsymbol{x}^t|g_k,\theta^i)P(g_k|\theta^i)}{\sum_{\acute{k}} P(\boldsymbol{x}^t|g_{\acute{k}},\theta^i)P(g_{\acute{k}}|\theta^i)} = \frac{|\boldsymbol{\Sigma}_k|^{-1/2}\exp\left(-\frac{1}{2}(\boldsymbol{x}^t-\boldsymbol{\mu}_k)^{\top}\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}^t-\boldsymbol{\mu}_k)\right)\pi_k}{\sum_{\acute{k}}|\boldsymbol{\Sigma}_{\acute{k}}|^{-1/2}\exp\left(-\frac{1}{2}(\boldsymbol{x}^t-\boldsymbol{\mu}_{\acute{k}})^{\top}\boldsymbol{\Sigma}_{\acute{k}}^{-1}(\boldsymbol{x}^t-\boldsymbol{\mu}_{\acute{k}})\right)\pi_{\acute{k}}} \equiv w_k^t$$

- Logarithm of a sum
  - Example
    - $e^{-1001} + e^{-1002}$
    - $l_1 = \log e^{-1001} = -1001$
    - $l_2 = \log e^{-1002} = -1002$
    - $$l_3 = \log\left(e^{l_1} + e^{l_2}\right) = \log\left(\underbrace{e^{-1001}}_{e^{-1001}e^{0}} + \underbrace{e^{-1002}}_{e^{-1001}e^{-1}}\right) = \log\left(e^{-1001}(e^{0} + e^{-1})\right)$$
    $$= \underbrace{\log e^{-1001}}_{-1001} + \log(e^{0} + e^{-1}) = l_1 + \log\left(1 + e^{l_2 - l_1}\right)$$
  - $$\log \sum_i p_i = \log(p_1 + p_2 + \cdots + p_n) = \log p_1\left(1 + \frac{p_2}{p_1} + \frac{p_3}{p_1} + \cdots + \frac{p_n}{p_1}\right)$$
    $$= \log p_1 + \log\left(1 + e^{\log\frac{p_2}{p_1}} + e^{\log\frac{p_3}{p_1}} + \cdots + e^{\log\frac{p_n}{p_1}}\right)$$
    $$= \log p_1 + \log\left(1 + e^{\log p_2 - \log p_1} + e^{\log p_3 - \log p_1} + \cdots + e^{\log p_n - \log p_1}\right)$$
    $$= l_1 + \log\left(1 + e^{l_2 - l_1} + e^{l_3 - l_1} + \cdots + e^{l_3 - l_1}\right) \qquad ; l_i \equiv \log p_i$$