$t_n = y(x_n, w) + \varepsilon$

$E(t/x)$

MMSE

$\mathbf{y} = \mathbf{W}\mathbf{f}$

$\mathbf{f} = [f_1(\mathbf{x})\ f_2(\mathbf{x}) \cdots f_r(\mathbf{x})]^T$

$$\mathbf{W} = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1r} \\ w_{21} & w_{22} & \cdots & w_{2r} \\ \cdots & \cdots & \cdots & \cdots \\ w_{m1} & w_{m2} & \cdots & w_{mr} \end{pmatrix}$$

$|\mathbf{T} - \mathbf{F}\mathbf{W}^T|^2$

$\mathbf{W}$,$\mathbf{F}$ and $\mathbf{T}$ are $r \times m$, $d \times m$ and $d \times n$

$\widehat{\mathbf{W}}^T = (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T\mathbf{T}$

Least square solution

i^{th} row of the matrix $\mathbf{T}$ in $\mathbf{T} = \mathbf{F}\mathbf{W}^T + \mathbf{E}$ Gaussian distributed with mean vector $\mathbf{f_i}\mathbf{W^T}$ co-variance matrix $\mathbf{I}\frac{1}{\beta}$

$\mathbf{f_i}$ is the i^{th} row of the matrix $\mathbf{F}$

likelihood function of the $\mathbf{T}$ given $\mathbf{W}$

$K\Pi_{i=1}^{i=d} e^{-|\mathbf{t_i} - \mathbf{f_i}\mathbf{W^T}|^2}$

$\widehat{\mathbf{W}}^T = (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T\mathbf{T}$

Likelihood solution

Regularization solution

$||\mathbf{T} - \mathbf{F}\mathbf{W}^T||_2^2 + \lambda||\mathbf{W}||_2^2$

$\widehat{\mathbf{W}}^T = (\mathbf{F}^T\mathbf{F} + \lambda\mathbf{I})^{-1}\mathbf{F}^T\mathbf{T}$

Note: The Bayes solution is identical with that of the Regularization solution if $\lambda = \frac{\alpha}{\beta}$

Bayes solution

The columns of the matrix W are Gaussian distributed with mean vector 0 and co-variance matrix $\frac{1}{\alpha}I$

posterior density function $p(\mathbf{W}/T)$

multi-variate Guassian distributed

$$\mathbf{M} = (\alpha I + \beta F^T F)^{-1}\beta F^T T = (F^T F + \tfrac{\alpha}{\beta} I)^{-1}(F^T)T$$

Sequential learning

$$W(t+1) = W(t) + \eta f(\mathbf{x_k})(\mathbf{t_k} - \mathbf{f}(\mathbf{x_k})^{\mathbf{T}}\mathbf{W})$$

where, $\mathbf{t_k}$ is the k^{th} row of the matrix T with size $1 \times n$.

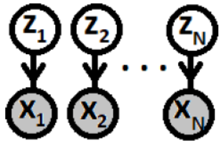
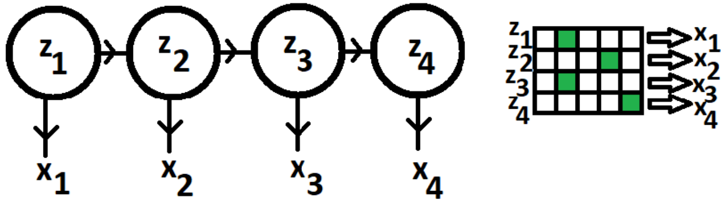
The size of the vector $f(x_k)$ is $M \times 1$

1. Let $\mathbf{x_1}, \mathbf{x_2}, \cdots \mathbf{x_d}$ are the training vectors.
2. Initialize the weight matrix $W(1)$ and initialize the value for $iter = 1$ and $k = 1$
3. Update the weight matrix $W(iter+1) = W(iter) + \eta f(\mathbf{x_k})(\mathbf{t_k} - \mathbf{f}(\mathbf{x_k})^{\mathbf{T}}\mathbf{W})$
4. Update $iter = iter + 1$ and $k = k + 1$.
5. Repeat (3) until $iter = d$.If $iter = d$,reset $iter = 1$ and $k = 1$. This is one epoch.
6. Compute the Sum Squared Error $||T - FW||^2$, using the latest updated matrix $\mathbf{W}$ after one epoch.
7. Repeat 3 and 4 until for number of epochs equal to N.
8. Declare the latest $\mathbf{W}$ as the solution.

Kernel smoothing method

$$\sum_{k=1}^{k=d} k(\mathbf{x}, \mathbf{x_k})t_k$$

1. Let the collected input data $\mathbf{x_1}, \cdots \mathbf{x_{100}}$ and the corresponding output data are $\mathbf{t_1}, \cdots \mathbf{t_{100}}$. These 100 data are divided into training and validation data as follows.
2. Training data:Input vectors $\mathbf{x_1}$, $\mathbf{x_2}$, $\cdots$ $\mathbf{x_{70}}$ and the corresponding target vector $\mathbf{t_1}$, $\mathbf{t_2}$,$\cdots$ $\mathbf{t_{70}}$ and the validation data:Input vectors $\mathbf{x_{71}}$,$\mathbf{x_{72}}$,$\cdots$ $\mathbf{x_{100}}$ and the corresponding target vectors $\mathbf{t_{71}}$,$\mathbf{t_{72}}$,$\cdots$ $\mathbf{t_{100}}$
3. Compute $\mathbf{t_u^{est}} = \sum_{k=1}^{k=70} k(\mathbf{x_k}, \mathbf{x_u})\mathbf{t_k}$ using the particular tuning parameter σ, (which is used in the kernel function) for $u = 71 \cdots 100$.
4. Compute the sum squared error as $SSE = \sum_{m=71}^{m=100}(t_m - t_m^{est})^2$
5. Repeat 3 and 4 for various values of σ and identify the σ that gives the minimum value of SSE.
6. Use the tuning parameter to model the linear regresssion model as $y(\mathbf{x}) = \sum_{k=1}^{k=70} k(\mathbf{x_k}, \mathbf{x})\mathbf{t_k}$.

Combinational Model (GMM)	Sequential Model (HMM)
Training set of the individual data are treated as independent $x_1 \; x_2 \; x_3 \; \cdots \; x_N$ The generating probability is given as the following $p(\mathbf{x}) = \sum_{k=1}^{k=r} \pi_k p(\mathbf{x}/\mu_k, C_k)$	$x_{r1} \; x_{r2} \; \ldots x_{rN}$ are the frames belonging to the r^{th} data The generating probability of the vector x (arbitrary sequential data) is given as the following $p(\mathbf{x}/\theta) = \sum_{\mathbf{z}} p(\mathbf{zx}/\theta)$
The unknown parameters are estimated by maximizing the generating probability $p(\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_N)$ using Expectation-Maximization algorithm $log(p(\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_N)) = \sum_{n=1}^{n=N} log(\sum_{k=1}^{k=r} \pi_k p(\mathbf{x}_n/\mu_k C_k))$	The unknown parameters are estimated by maximizing the following. $\sum_{\mathbf{z}} p(\mathbf{z}/\mathbf{x}\theta_{old}) ln(p(\mathbf{zx}/\theta))$
	
Expectation $r_m(\mathbf{x}_n) = \frac{\pi_m p(\mathbf{x}_n/\mu_m C_k)}{(\sum_{k=1}^{k=N} \pi_k p(\mathbf{x}_n/\mu_k C_k))}$ **Maximization** $\mu_m = \frac{\sum_{n=1}^{n=N} r_m(\mathbf{x}_n)\mathbf{x}_n}{\sum_{n=1}^{n=N} r_m(\mathbf{x}_n)}$ $C_m = \frac{\sum_{n=1}^{n=N} r_m(\mathbf{x}_n)(\mathbf{x}_n - \mu_m)((\mathbf{x}_n - \mu_m))^T}{\sum_{n=1}^{n=N} r_m(\mathbf{x}_n)}$ $\pi_m = \frac{r_m(\mathbf{x}_n)}{\sum_{n=1}^{n=N} r_m(\mathbf{x}_n)}$	**Expectation** $\alpha(\mathbf{z}_n) = p(\mathbf{x}_n/\mathbf{z}_n) \sum_{z_{n-1}} \alpha(\mathbf{z}_{n-1}) p(\mathbf{z}_n/\mathbf{z}_{n-1})$ $\beta(\mathbf{z}_n) = \sum_{\mathbf{z}_{n+1}} p(\mathbf{x}_{n+1}/\mathbf{z}_{n+1}) \beta(\mathbf{z}_{n+1}) p(\mathbf{z}_{n+1}/\mathbf{z}_n)$ $\gamma_{\mathbf{z}_n} = \frac{\alpha(\mathbf{z}_n)\beta(\mathbf{z}_n)}{p(\mathbf{X})}$ $\varepsilon(n, n-1) = p(\mathbf{z}_n, \mathbf{z}_{n-1}/\mathbf{X}) = \frac{\alpha(\mathbf{z}_{n-1}) p(\mathbf{x}_n/\mathbf{z}_n) \beta(\mathbf{z}_n) p(\mathbf{z}_n/\mathbf{z}_{n-1})}{p(\mathbf{X})}$ **Maximization** $\pi_l = \frac{\gamma_{z_{1,l}}}{\sum_{r=1}^{r=K} \gamma_{z_{1,r}}}$ $A_{i,j} = \frac{\sum_{n=2}^{n=N} \varepsilon(n,i,n-1,j)}{\sum_{i=1}^{i=K} \sum_{n=2}^{n=N} \varepsilon(n,i,n-1,j)}$ $\mu_k = \frac{\sum_{n=1}^{n=N} \gamma_{z_{n,k}} X_n}{\sum_{n=1}^{n=N} \gamma_{z_{n,k}}}$ $\sum_k = \frac{\sum_{n=1}^{n=N} \gamma_{z_{n,k}} (\mathbf{x}_n - \mu_k)((\mathbf{x}_n - \mu_k))^T}{\sum_{n=1}^{n=N} \gamma_{z_{n,k}}}$