3) To update Σ, we will prefer working with $\Lambda := \Sigma^{-1}$

We assume Σ is invertible and to obtain an update rule for Λ we will differentiate the objective function in the M-step w.r.t Λ

$$\sum_{i=1}^{N}\sum_{m=1}^{M} w_m^{(i)} \ln N(\underline{x}^{(i)} \mid \underline{\mu}_m, \Lambda^{-1}) + \ln \pi_m)$$

$$= \sum_{i=1}^{N}\sum_{m=1}^{M} w_m^{(i)} \ln \left[\frac{1}{(2\pi)^{p/2} |\Lambda^{-1}|^{1/2}} \exp\left(-\frac{1}{2} (\underline{x}^{(i)} - \underline{\mu}_m)^T \Lambda (\underline{x}^{(i)} - \underline{\mu}_m)\right)\right]$$

$$+ \sum_{i=1}^{N}\sum_{m=1}^{M} \ln \pi_m$$

$$= - \sum_{i=1}^{N}\sum_{m=1}^{M} w_m^{(i)} \left[\frac{p}{2} \ln 2\pi - \frac{1}{2} \ln |\Lambda| + \frac{1}{2} (\underline{x}^{(i)} - \underline{\mu}_m)^T \Lambda (\underline{x}^{(i)} - \underline{\mu}_m) \right]$$

$$+ \sum_{i=1}^{N}\sum_{m=1}^{M} \ln \pi_m$$

$$= - \sum_{i=1}^{N}\sum_{m=1}^{M} \frac{w_m^{(i)} (\underline{x}^{(i)} - \underline{\mu}_m)^T \Lambda (\underline{x}^{(i)} - \underline{\mu}_m)}{2} + \sum_{i=1}^{N}\sum_{m=1}^{M} \frac{w_m^{(i)} \ln |\Lambda|}{2}$$

$$- \sum_{i=1}^{N}\sum_{m=1}^{M} \frac{w_m^{(i)} p \ln 2\pi}{2} + N \sum_{m=1}^{M} \ln \pi_m$$

$$\propto - \sum_{i=1}^{N}\sum_{m=1}^{M} \frac{w_m^{(i)} (\underline{x}^{(i)} - \underline{\mu}_m)^T \Lambda (\underline{x}^{(i)} - \underline{\mu}_m)}{2} + \sum_{i=1}^{N}\sum_{m=1}^{M} \frac{w_m^{(i)} \ln |\Lambda|}{2}$$

Objective function having Λ

$$\alpha - \sum_{i=1}^{N}\sum_{m=1}^{M} \frac{w_m^{(i)} (\underline{x}^{(i)} - \underline{\mu}_m)^T \Lambda (\underline{x}^{(i)} - \underline{\mu}_m)}{2} + \sum_{i=1}^{N}\sum_{m=1}^{M} \frac{w_m^{(i)} \ln|\Lambda|}{2}$$

$$\alpha - \sum_{i=1}^{N}\sum_{m=1}^{M} \frac{w_m^{(i)} (\underline{x}^{(i)} - \underline{\mu}_m)^T \Lambda (\underline{x}^{(i)} - \underline{\mu}_m)}{2} + \sum_{i=1}^{N}\underbrace{\sum_{m=1}^{M} w_m^{(i)}}_{\text{sum to } 1} \frac{\overbrace{\ln|\Lambda|}^{\text{independent of } i}}{2}$$

$$\alpha - \sum_{i=1}^{N}\sum_{m=1}^{M} \frac{w_m^{(i)} (\underline{x}^{(i)} - \underline{\mu}_m)^T \Lambda (\underline{x}^{(i)} - \underline{\mu}_m)}{2} + \frac{N \ln|\Lambda|}{2}$$

Setting the gradient of the objective function to zero & solving for Λ

Use the following identities:

$$\frac{\partial\, \underline{a}^T \underline{\underline{X}}\, \underline{b}}{\partial \underline{\underline{X}}} = \underline{b}\, \underline{a}^T$$

$$\frac{\partial \ln|\underline{\underline{X}}|}{\partial \underline{\underline{X}}} = \underline{\underline{X}}^{-T}$$

$$-\sum_{i=1}^{N}\sum_{m=1}^{M} w_m^{(i)} (\underline{x}^{(i)} - \underline{\mu}_m)(\underline{x}^{(i)} - \underline{\mu}_m)^T + N\Lambda^{-1} = 0$$

$$\Rightarrow \quad \hat{\Lambda} = \left(\frac{\sum_{i=1}^{N}\sum_{m=1}^{M} w_m^{(i)} (\underline{x}^{(i)} - \underline{\mu}_m)(\underline{x}^{(i)} - \underline{\mu}_m)^T}{N} \right)^{-1}$$

Therefore, update rule of Σ is:

$$\hat{\Sigma} \leftarrow \frac{\sum_{i=1}^{N}\sum_{m=1}^{M} w_m^{(i)} (\underline{x}^{(i)} - \underline{\mu}_m)(\underline{x}^{(i)} - \underline{\mu}_m)^T}{N}$$

2) Bias-variance decomposition

$$\mathbb{E}[(y - \hat{y})^2] = (\text{Bias of the model})^2 + (\text{Variance of the model})^2 + \text{Irredicible error due to noise}$$

noisy measurement (y)

prediction from model ($\hat{y}$)

Let the mean prediction and variance of each member in the bagging ensemble be μ and σ^2 respectively.

$$y_{bag} = \frac{1}{B} \sum_{j=1}^{B} y_j$$

y_j ← prediction from the jth member

If the true I/O relation is given by: $y = f_0(\underline{x}) + \epsilon$

then bias of bagged model is:

$$\left(\mathbb{E}[y_{bag}] - f_0(\underline{x})\right)^2 = \left(\mathbb{E}\left[\frac{1}{B}\sum_{j=1}^{B} y_j\right] - f_0(\underline{x})\right)^2$$

$$= \left(\frac{1}{B}\sum_{j=1}^{B} \underbrace{\mathbb{E}(y_j)}_{\text{mean prediction of jth ensemble member}} - f_0(\underline{x})\right)^2$$

All ensemble members are assumed to have the same mean prediction

$$= \left(\frac{1}{B} \cdot B\mu - f_0(\underline{x})\right)^2$$

$$= (\mu - f_0(\underline{x}))^2$$

So the bias of the bagging ensemble remains unchanged

Variance of bagged model

mean prediction of bagged model

$$\mathrm{Var}(y_{bag}) = \mathbb{E}_T\left[\left(y_{bag} - \mu\right)^2\right]$$

$$= \mathbb{E}_T\left[\left(\frac{1}{B}\sum_{j=1}^{B} \underbrace{(y_j - \mu)}_{z_j}\right)^2\right]$$

$$= \frac{1}{B^2}\sum_{j=1}^{B} \mathrm{Var}(y_j) + \frac{1}{B^2}\sum_{j=1}^{B}\sum_{i \neq j} \mathrm{Cov}(y_i, y_j)$$

Variance of each ensemble member:

$$\mathrm{Var}(y_j) = \mathbb{E}_T\left[(y_j - \mu)^2\right] = \sigma^2$$

$$\underset{i \neq j}{\mathrm{Cov}(y_i\, y_j)} = \mathbb{E}_T\left[(y_i - \mu)(y_j - \mu)\right]$$

$$= \rho\sigma^2$$

correlation

$$= \frac{1}{B^2}\cdot(B\sigma^2) + \frac{1}{B^2}B\cdot(B-1)\,\rho\sigma^2$$

$$= \frac{\sigma^2}{B} + \left(1 - \frac{1}{B}\right)\rho\sigma^2 = \left(\frac{1-\rho}{B}\right)\sigma^2 + \rho\sigma^2$$

Since the correlation $\rho \leq 1$, we see that

$$\rho\sigma^2\left(1 - \frac{1}{B}\right) \leq \sigma^2\left(1 - \frac{1}{B}\right)$$

$$\therefore \quad \frac{\sigma^2}{B} + \rho\sigma^2\left(1 - \frac{1}{B}\right) \leq \sigma^2$$

individual model variance