CWVSmix: Critical Window Variable Selection for Mixtures

Statistical Model

$$Y_i | p_i \overset{\text{ind}}{\sim} \text{Bernoulli}\left(p_i\right), \ i = 1, ..., n;$$

$$\ln\left(\frac{p_i}{1 - p_i}\right) = \mathbf{x}_i^{\text{T}} \boldsymbol{\beta} + \sum_{t=1}^{m_i} \left[\sum_{j=1}^{q} \lambda_j\left(t\right) z_{ij}\left(t\right) + \sum_{j=1}^{q-1} \sum_{k=j+1}^{q} \widetilde{\lambda}_{jk}\left(t\right) z_{ij}\left(t\right) z_{ik}\left(t\right) \right] \alpha\left(t\right).$$

Time-Varying Weights

$$\lambda_j\left(t\right) = \frac{\max\left\{\lambda_j^*\left(t\right), 0\right\}}{d\left(t\right)},$$

$$\widetilde{\lambda}_{jk}\left(t\right) = \frac{\max\left\{\widetilde{\lambda}_{jk}^*\left(t\right), 0\right\} 1\left\{\lambda_j^*\left(t\right) > 0\right\} 1\left\{\lambda_k^*\left(t\right) > 0\right\}}{d\left(t\right)},$$

$$t = 1, ..., m, \ j = 1, ..., q - 1, \ k = j + 1, ..., q;$$

$$d\left(t\right) = \sum_{j=1}^{q} \max\left\{\lambda_j^*\left(t\right), 0\right\} + \sum_{j=1}^{q-1} \sum_{k=j+1}^{q} \max\left\{\widetilde{\lambda}_{jk}^*\left(t\right), 0\right\} 1\left\{\lambda_j^*\left(t\right) > 0\right\} 1\left\{\lambda_k^*\left(t\right) > 0\right\};$$

$$\boldsymbol{\lambda}^* = \left\{\boldsymbol{\lambda}^*\left(1\right)^{\text{T}}, \ldots, \boldsymbol{\lambda}^*\left(m\right)^{\text{T}}\right\}^{\text{T}} | \phi_\lambda \sim \text{MVN}\left\{\mathbf{0}_{mq(q+1)/2}, \ \Sigma\left(\phi_\lambda\right) \otimes I_{q(q+1)/2}\right\}.$$

- $m = \max\left\{m_i : i = 1, ..., n\right\};$
- $\boldsymbol{\lambda}^*\left(t\right) = \left\{\lambda_1^*\left(t\right), \ldots, \lambda_q^*\left(t\right), \widetilde{\lambda}_{12}^*\left(t\right), \ldots, \widetilde{\lambda}_{q-1,q}^*\left(t\right)\right\}^{\text{T}};$
- $\mathbf{0}_{mq(q+1)/2}$: Length $mq(q+1)/2$ vector with each entry equal to zero;
- $I_{q(q+1)/2}$: $q(q+1)/2$ by $q(q+1)/2$ identity matrix.

Mixture Risk

$$\alpha\left(t\right) = \theta\left(t\right) \gamma\left(t\right), \ t = 1, ..., m;$$

$$\gamma\left(t\right) | \pi\left(t\right) \overset{\text{ind}}{\sim} \text{Bernoulli}\left\{\pi\left(t\right)\right\}, \ \Phi^{-1}\left\{\pi\left(t\right)\right\} = \eta\left(t\right);$$

$$\begin{bmatrix} \theta\left(t\right) \\ \eta\left(t\right) \end{bmatrix} = A \begin{bmatrix} \delta_1\left(t\right) \\ \delta_2\left(t\right) \end{bmatrix}, \ A = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix};$$

$$\boldsymbol{\delta}_j = \left\{\delta_j\left(1\right), ..., \delta_j\left(m\right)\right\}^{\text{T}} | \phi_j \overset{\text{ind}}{\sim} \text{MVN}\left\{\mathbf{0}_m, \Sigma\left(\phi_j\right)\right\}, \ j = 1, 2.$$

Prior Information

$\beta_j \overset{\text{iid}}{\sim} \text{N}\left(0, \sigma_\beta^2\right), \ j = 1, ..., p;$

- p: Length of $\mathbf{x}_{ij}$ vector (same for all i, j);
- Default setting: $\sigma_\beta^2 = 10,000$.

$\ln\left(A_{11}\right), \ln\left(A_{22}\right), A_{21} \overset{\text{iid}}{\sim} \text{N}\left(0, \sigma_A^2\right);$

- Default setting: $\sigma_A^2 = 1.00$.

$\phi_\lambda, \phi_j \overset{\text{iid}}{\sim} \text{Gamma}\left(\alpha_{\phi_j}, \beta_{\phi_j}\right), \ j = 1, 2;$

- Default setting: $\alpha_{\phi_j} = 1.00, \ \beta_{\phi_j} = 1.00, \ j = 1, 2$.

Default Initial Values

- $\beta_j = 0$ for all j;
- $\gamma\left(t\right) = 1$ for all t;
- $\delta_j\left(t\right) = 0$ for all j, t;
- $\phi_\lambda, \phi_j = -\ln\left(0.05\right)/\left(m - 1\right)$ for all j;
- $A_{jj} = 1$ for all j;
- $A_{21} = 0$.

Interactions

- interaction_indicator $= 0$:

$$\ln\left(\frac{p_i}{1 - p_i}\right) = \mathbf{x}_i^{\text{T}}\boldsymbol{\beta} + \sum_{t=1}^{m_i}\sum_{j=1}^{q}\lambda_j\left(t\right)\text{z}_{ij}\left(t\right)\alpha\left(t\right);$$

- interaction_indicator $= 1$: Full model with interactions as detailed above.

Alternate Likelihood

$Y_i|\mu_i, \sigma_\epsilon^2 \overset{\text{ind}}{\sim} \text{Normal}\left(\mu_i, \sigma_\epsilon^2\right), \ i = 1, ..., n;$

$\mu_i = \mathbf{x}_i^{\text{T}}\boldsymbol{\beta} + \sum_{t=1}^{m_i}\left[\sum_{j=1}^{q}\lambda_j\left(t\right)\text{z}_{ij}\left(t\right) + \sum_{j=1}^{q-1}\sum_{k=j+1}^{q}\widetilde{\lambda}_{jk}\left(t\right)\text{z}_{ij}\left(t\right)\text{z}_{ik}\left(t\right)\right]\alpha\left(t\right).$

- $\sigma_\epsilon^2 \sim \text{Inverse Gamma}\left(a_{\sigma_\epsilon^2}, b_{\sigma_\epsilon^2}\right);$
- Default setting: $a_{\sigma_\epsilon^2} = 0.01, \ b_{\sigma_\epsilon^2} = 0.01;$
- Default initial value: $\sigma_\epsilon^2 = 1.00$.
- likelihood_indicator $= 0$: Bernoulli;
- likelihood_indicator $= 1$: Gaussian.