
Algorithm 4: Supervised learning cut level algorithm

input : Training set $\mathbf{T} = \mathbf{Z}_{S_1}^{train}$ and test set $\mathbf{U} = \mathbf{Z}_{S_1}^{test}$
output: Matrix $\tilde{\mathbf{X}}^{(s^*)}$ of aggregated-SNP at best cut level s^*

hierarchy \leftarrow Constrained-HAC on \mathbf{T}
cutlevel \leftarrow Initialize levels where to cut hierarchy
for $s \leftarrow$ Sequence(cutlevel) do
 $\tilde{\mathbf{T}}^s \leftarrow$ Aggregating(\mathbf{T} , hierarchy, cutlevel[s]);
 $\tilde{\mathbf{U}}^s \leftarrow$ Aggregating(\mathbf{U} , hierarchy, cutlevel[s]);
 ridgecoef \leftarrow RidgeRegression($\mathbf{y}_{S_1}^{train} \sim \tilde{\mathbf{T}}$);
 $\mathbf{y}_{S_1}^{pred} \leftarrow$ Predict($\tilde{\mathbf{U}}$, ridgecoef);
 AUC[s] \leftarrow ROC($\mathbf{y}_{S_1}^{test}, \mathbf{y}_{S_1}^{pred}$);
end
 $s^* \leftarrow$ Which(cutlevel, Max(AUC));
 $\tilde{\mathbf{X}}^{(s^*)} \leftarrow$ Aggregating(\mathbf{Z} , hierarchy, bestlevel);
