Algorithm 1 Training and Prediction for MTGB py

Require: Training data X, targets y, task info $task\_info$

Initialize model parameters:

$learning\_rate(\eta) \leftarrow 0.1$, $max\_depth \leftarrow 1$

Initialize base estimator with dummy values

Training Phase:

for i in $n\_estimators$ **do**

Compute negative gradient:

$$\text{neg\_grad} \leftarrow y - \hat{y}$$

if $i = 0$ **then**

Initialize ensemble prediction:

$$p_{\text{meta}} \leftarrow p_{\text{meta}} + (1 - \sigma(\theta)) \times p_{\text{non\_out}} + \sigma(\theta) \times p_{\text{out}} + p_{\text{task}}$$

else if $i \leq n\_common\_estimators$ **then**

Update meta ensemble prediction:

$$p_{\text{meta}} \leftarrow p_{\text{meta}} + \eta \times \text{tree(X, neg\_grad)}$$

else if $i \leq n\_mid\_estimators$ **then**

Update ensemble prediction for outlier and non-outlier block:

$$p_{\text{non\_out}}, p_{\text{out}} \leftarrow p_{\text{meta}} + (1 - \sigma(\theta)) \times p_{\text{non\_out}} + \sigma(\theta) \times p_{\text{out}} + p_{\text{task}}$$

Compute gradients for outlier and non-outlier:

$$\text{neg\_grad\_outlier} \leftarrow \text{neg\_grad} \times \sigma(\theta)$$

$$\text{neg\_grad\_non\_outlier} \leftarrow \text{neg\_grad} \times (1 - \sigma(\theta))$$

Update outlier estimator:

$$p_{\text{out}} \leftarrow p_{\text{out}} + \eta \times \text{tree(X, neg\_grad\_outlier)}$$

Update non-outlier estimator:

$$p_{\text{non\_out}} \leftarrow p_{\text{non\_out}} + \eta \times \text{tree(X, neg\_grad\_non\_outlier)}$$

Optimize task-specific parameter θ:

$$\theta \leftarrow \theta - \eta \frac{\partial L}{\partial \theta}$$

Gradient of Loss w.r.t. θ (for each task):

$$\frac{\partial L}{\partial \theta} = \sigma(\theta) \times (1 - \sigma(\theta)) \times (p_{\text{out}} - p_{\text{non\_out}})$$

else

Update task-specific prediction:

Update ensemble prediction for task-specific block:

$$p_{\text{task}} \leftarrow p_{\text{meta}} + (1 - \sigma(\theta)) \times p_{\text{non\_out}} + \sigma(\theta) \times p_{\text{out}} + p_{\text{task}}$$

$$p_{\text{task}} \leftarrow p_{\text{task}} + \eta \times \text{tree(X, neg\_grad)}$$

end if

end for

Prediction Phase:

Compute meta, outlier, non-outlier, and task predictions:

$$p_{\text{meta}}, p_{\text{out}}, p_{\text{non\_out}}, p_{\text{task}} \leftarrow \text{initial predictions from models}$$