

```

1 foreach TBC model,  $m$  do
2   Specify training and holdout data,  $T_m$  and  $H_m$ 
3   Fit probability model,  $f_{m,T}$ , using data from  $T_m$ 
4   Apply model to  $T_m$  and  $H_m$ :  $f_{m,T}(T_m)$  and  $f_{m,T}(H_m)$ 
5   foreach data set  $D \in \{T_m, H_m\}$  do
6     foreach cutoff  $c \in [500 \text{ point linear grid of } (0, 1)]$  do
7       Assign classifications:  $class_{m,T,c}(D) = \text{integer}(f_{m,T}(D) > c)$ 
8       Calculate MCC value between the model's response,  $y_D$ , and
        the latter classifications:
        
$$MCC_{m,T,c}(D) = \text{corr}(y_D, class_{m,T,c}(D))$$

9   foreach Bootstrap iteration,  $i$  do
10    Get bootstrapped samples  $T_{m,i}^*$  and  $H_{m,i}^*$ 
11    Fit bootstrapped probability model,  $f_{m,T,i}^*$ , using bootstrapped
        data from  $T_{m,i}^*$ 
12    Apply bootstrapped model to original data and bootstrapped
        samples:  $f_{m,T,i}^*(T_m)$ ,  $f_{m,T,i}^*(H_m)$ ,  $f_{m,T,i}^*(T_{m,i}^*)$ ,  $f_{m,T,i}^*(H_{m,i}^*)$ 
13    foreach data set  $D \in \{T_m, H_m, T_{m,i}^*, H_{m,i}^*\}$  do
14      foreach cutoff  $c \in [500 \text{ point linear grid of } (0, 1)]$  do
15        Assign bootstrapped classifications:
        
$$class_{m,T,i,c}^*(D) = \text{integer}(f_{m,T,i}^*(D) > c)$$

16        Calculate bootstrapped MCC value between the model's
        response,  $y_D$ , and the latter classifications:
        
$$MCC_{m,T,i,c}^*(D) = \text{corr}(y_D, class_{m,T,i,c}^*(D))$$

17    foreach original data  $D \in \{T_m, H_m\}$  and bootstrapped sample
         $D^* \in \{T_{m,i}^*, H_{m,i}^*\}$  do
18      foreach cutoff  $c \in [500 \text{ point linear grid of } (0, 1)]$  do
19        Calculate bootstrapped optimism for MCC value at cutoff  $c$ :
        
$$O_{m,T,c}(D) = \mathbf{E}_i[MCC_{m,T,i,c}^*(D^*) - MCC_{m,T,i,c}^*(D)]$$

20        Calculate optimism-corrected MCC value at cutoff  $c$ :
        
$$\widetilde{MCC}_{m,T,c}(D) = MCC_{m,T,c}(D) - O_{m,T,c}(D)$$

21        Average MCC values across bootstrapped iterations:
        
$$\overline{MCC}_{m,T,c}^*(D) = \mathbf{E}_i[MCC_{m,T,i,c}^*(D)]$$

22    Find optimal cutoff,  $c_{m,T}^*(D)$ , that maximizes mean of
        bootstrapped MCC values:  $c_{m,T}^*(D) = \max_c \overline{MCC}_{m,T,c}^*(D)$ 

```