

Solution 1:

- A) no resampling.
- B) resampling.
- C) Nested resampling.

Always everything, but with different roles (train, val, test)

Tune before training, rest does not matter, i.e. possible: BAC, BCA, CBA

First GE, then Tune, then Train (CBA):

GE:

Algorithm 1 GE: Outer loop (3-fold CV)

```
for  $k \in \{1, 2, 3\}$  do
  Inner loop (4-fold CV):
  for  $j \in \{1, 2, 3, 4\}$  do
    for each hyperparameter  $l$  do
       $f_{l,k,j}$  = train model on  $D_{\text{train}_{k,j}}$  with hyperparameter  $l$ 
       $ge_{l,k,j}$  = test error of  $f_{l,k,j}$  on  $D_{\text{test}_{k,j}}$ 
    end for
  end for
   $ge_{l,k}$  =  $\text{mean}_j ge_{l,k,j}$ 
   $l_k^*$  =  $\arg \min_l ge_{l,k}$  ▷ best hyperparameter combination in this fold of the outer loop
   $f_{l^*,k}$  = train model on  $D_{\text{train}_k}$ 
   $ge_k$  = test error of  $f_{l^*,k}$  on  $D_{\text{test}_k}$ 
end for
 $\hat{ge}$  =  $\text{mean}(ge_k)$  ▷ final estimate of generalization error
```

Tuning:

Algorithm 2 Tuning

```
for  $j \in \{1, 2, 3, 4\}$  do
  for each hyperparameter  $l$  do
     $f_{l,j}$  = train model on  $D_{\text{train}_j}$  with hyperparameter  $l$ 
     $ge_{l,j}$  = test error of  $f_{l,j}$  on  $D_{\text{test}_j}$ 
  end for
end for
 $ge_l$  =  $\text{mean}_j ge_{l,j}$ 
 $l^*$  =  $\arg \min_l ge_l$  ▷ best hyperparameter combination
```

Training:

Algorithm 3 Training

```
 $f^*$  = train model on  $D$  with hyperparameter  $l^*$ 
```

Total number of hyperparameter combinations: $5 + 4 \times 4 = 21$ (either NN or RF, and then grid of all hyperparameters). With this:

A) Final training: 1 model

B) Tuning of graph learner: $4 \times 21 = 84$ models (each hyperparameter in each fold)

C) Estimation GE: $3 \times 4 \times 21$ (Tuning whole graph per fold) + 3 (computing test error per fold) or $3 \times (4 \times 21 + 1) = 255$ models

Total = 340 models