GCM, LOCO Subsets selection

Chenghui Zheng

2024-08-13

(a) Linear Model with Independent Predictors

$$Y_1 \sim 1.5X_1 + 1.5X_2 + 2X_3 + 2X_4 + 2X_5 + 3X_6 + 4X_7 + 5X_8 + \epsilon$$

(b) Linear Model with Correlated Predictors

$$Y_2 \sim 1.5X_1 + 1.5X_2 + 2X_3 + 2X_4 + 2X_5 + 3X_6 + 4X_7 + 5X_8 + \epsilon$$

Where $X_1 \not\perp\!\!\!\perp X_2$ and $\text{cov}(X_1, X_2) = 0, 0.5, 0.75, 0.9$ respectively.

(c) Linear Model with Correlated Predictors and Different SNR

$$Y_3 \sim 1.5X_1 + 1.5X_2 + 2X_3 + 2X_4 + 2X_5 + 3X_6 + 4X_7 + 5X_8 + \epsilon$$

Where $\text{cov}(X_i, X_j) = \rho^{|i-j|}$ and $\epsilon \sim N(0, \sigma^2)$ with $\sigma^2 = 0.1, 0.5, 0.75, 2.1$.

(d) Non-linear Model

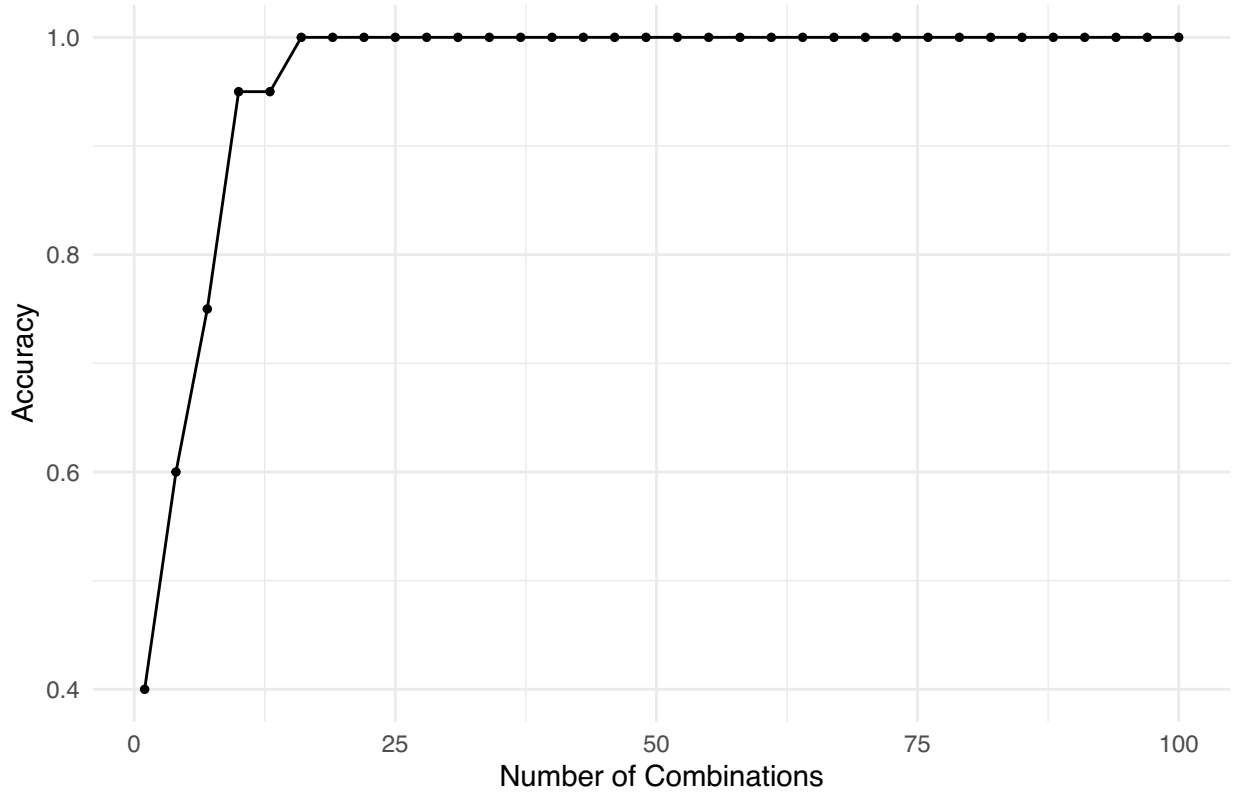
$$Y_4 \sim 2X_1^2 + 2\cos(4X_2) + \sin(X_3) + \exp(X_4/3) + 3X_5 + X_6^3 + 5X_7 + \max(0, X_8)$$

GCM Subsets selection

```
##      user  system elapsed
##    131.00    1.27  132.88
```

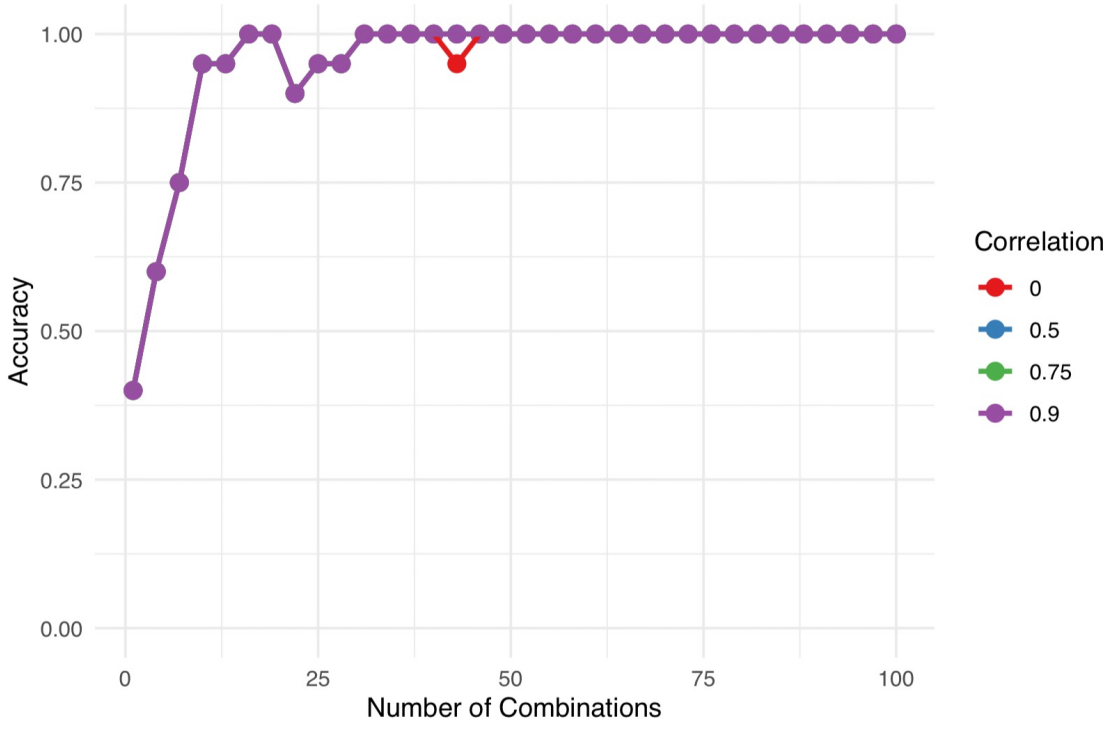
```
   user  system elapsed
 131.00    1.27  132.88
```

Case a):



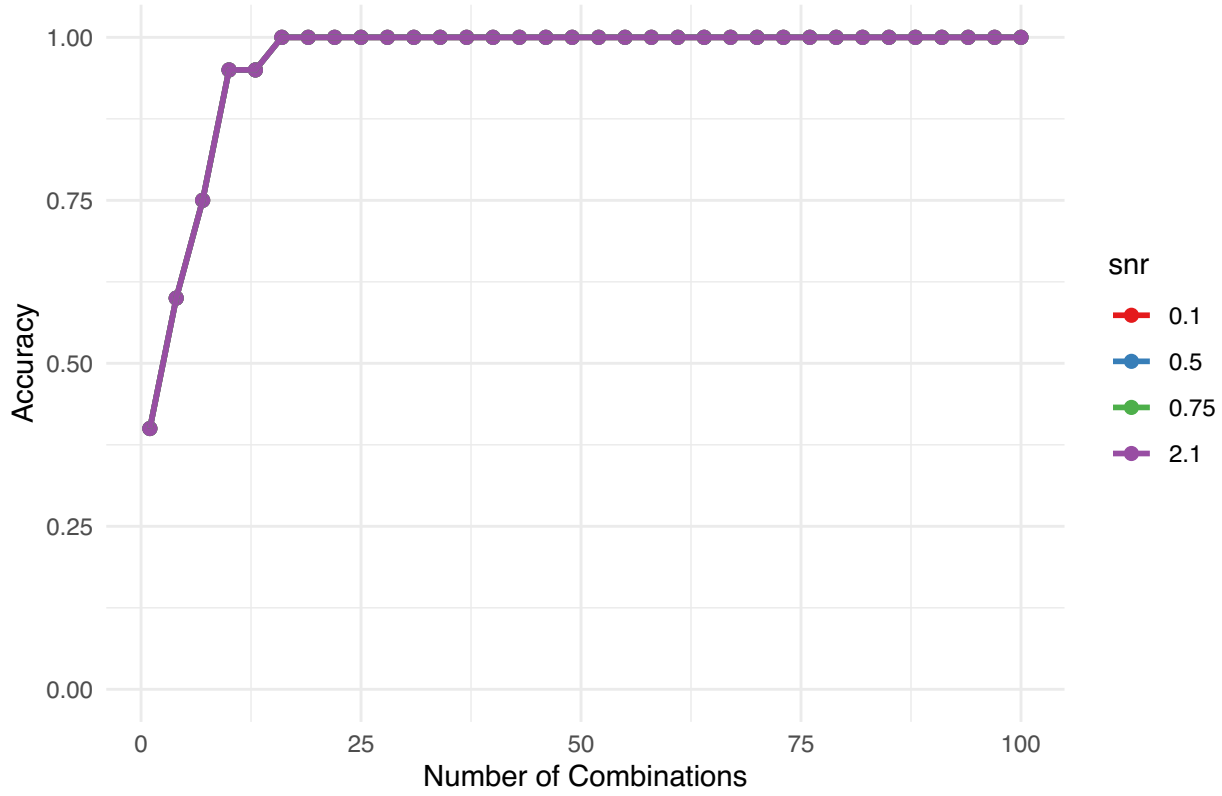

```
##    user  system elapsed
##  678.85    7.23  700.61
```

Case b):



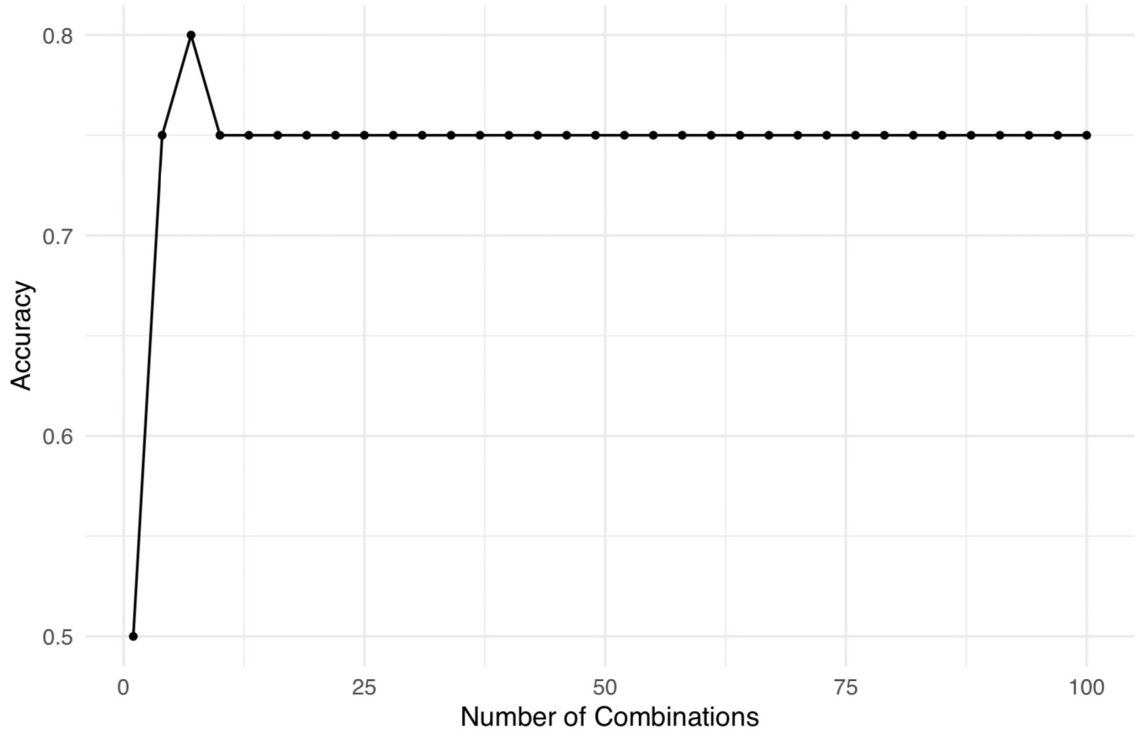
```
##    user  system elapsed
## 1323.74   13.22 1361.04
```

Case c):




```
##       user  system elapsed
##     347.26    1.97  353.15
```

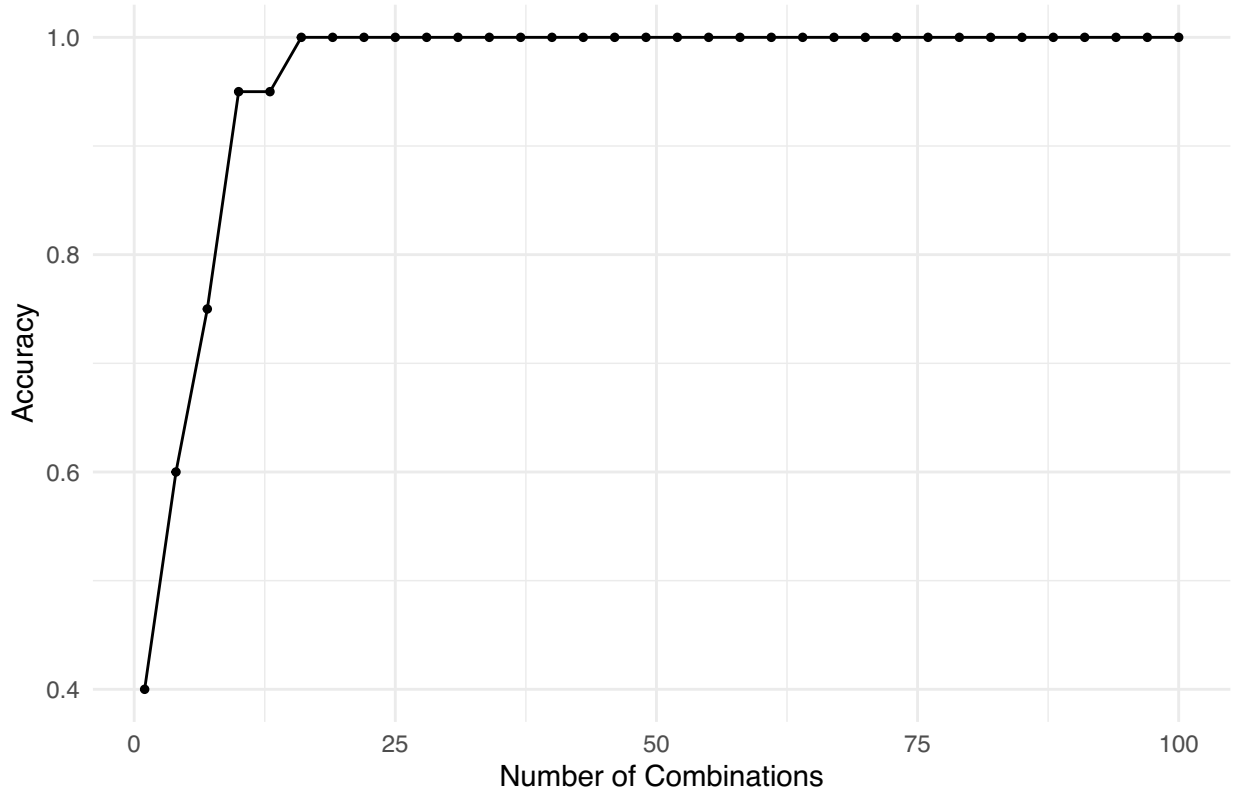
Case d):



Loco subset selection

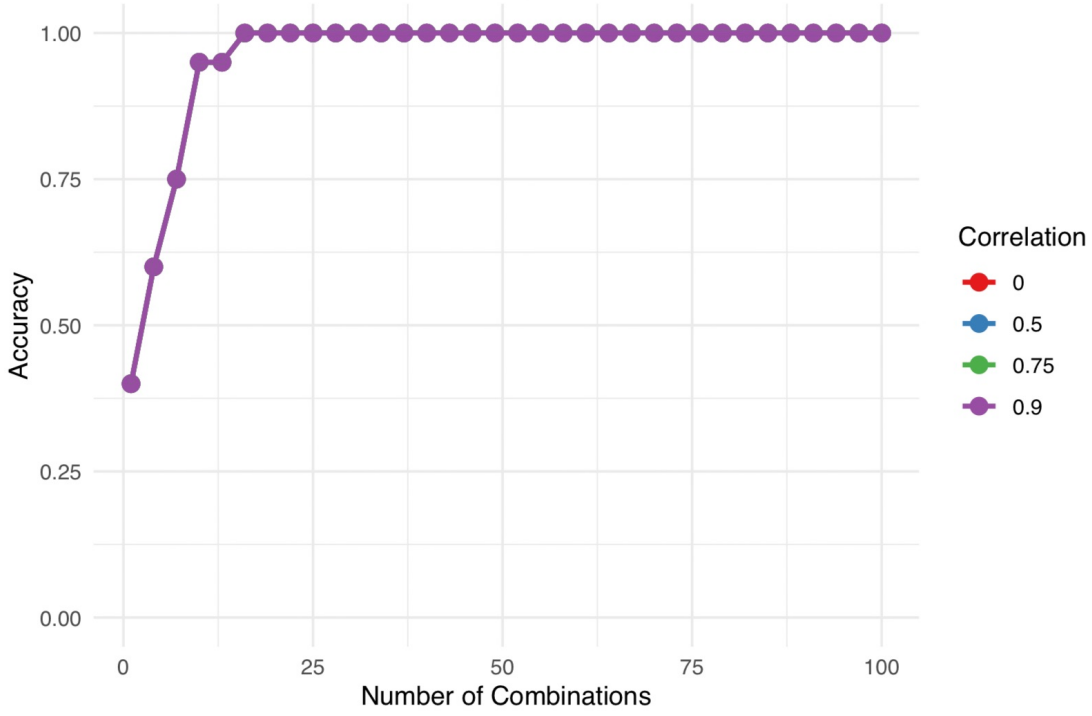
```
##    user  system elapsed
##   97.19    0.50   98.87
```

Case a):



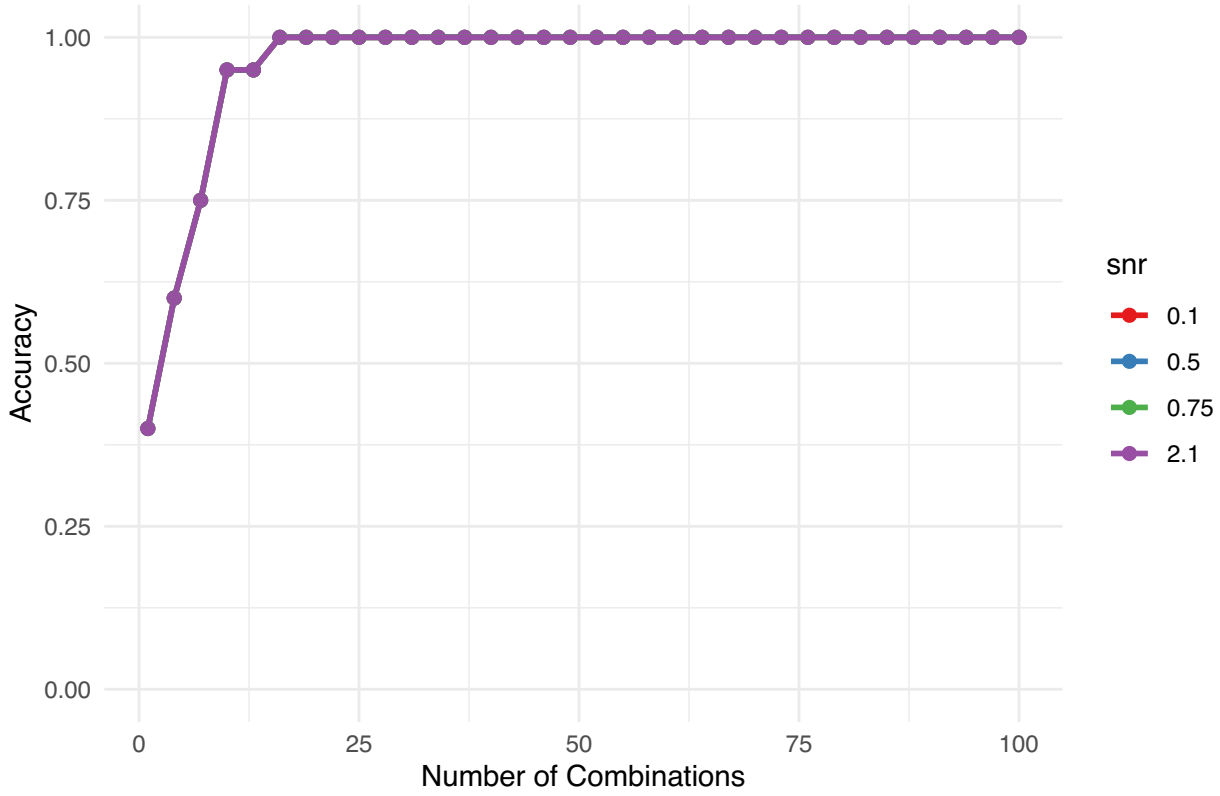

```
##    user  system elapsed
##  414.16    2.11  421.13
```

Case b):



```
##    user  system elapsed
##  813.85    3.72  826.95
```

Case c):




```
##    user  system elapsed
##  128.66    0.30  130.29
```

Case d):