

GCM, LOCO Subsets selection

Chenghui Zheng

2024-08-19

(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 $X_1 \not\perp X_2$ 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

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

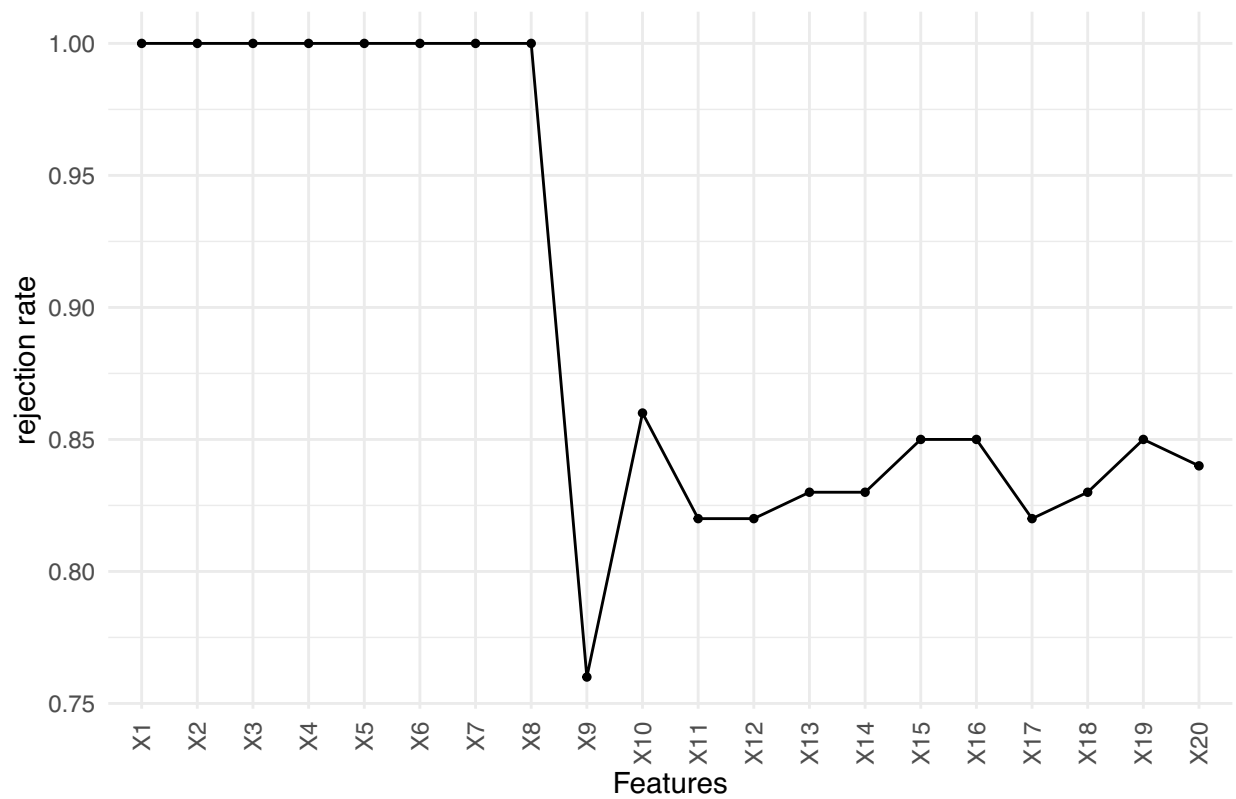
GCM Subsets selection(old T(n),100 simulations, each 500 instances)

```
##      user  system elapsed
## 137.06    1.41   140.71
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

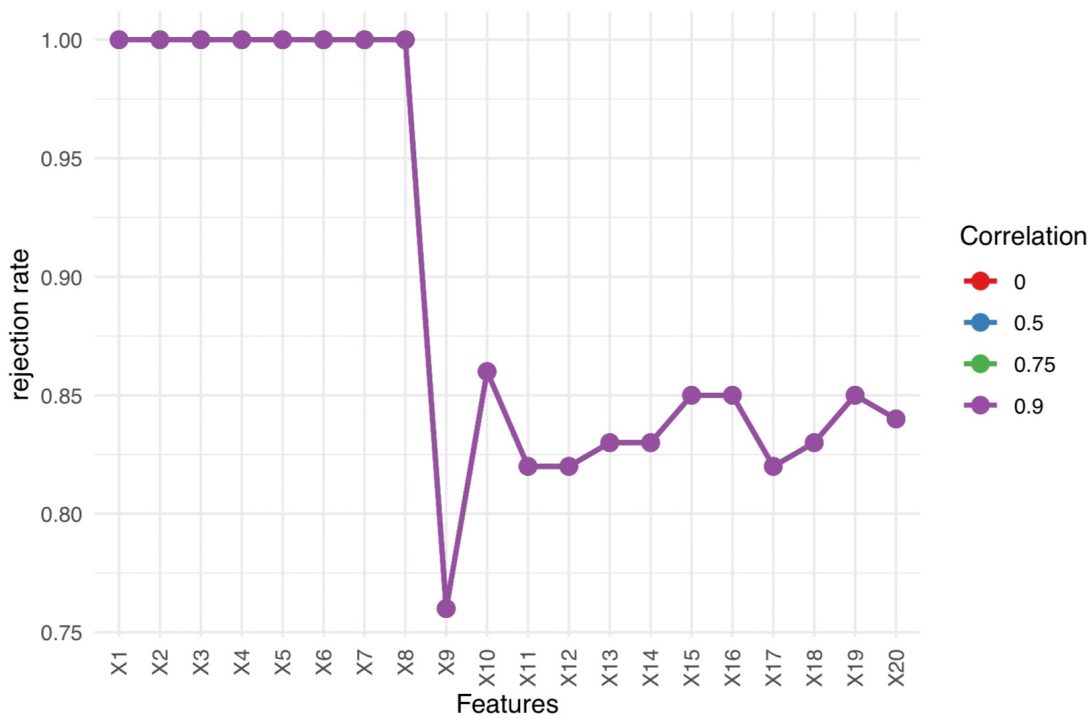
```
##      user  system elapsed
## 137.06    1.41  140.71
```

Case a):



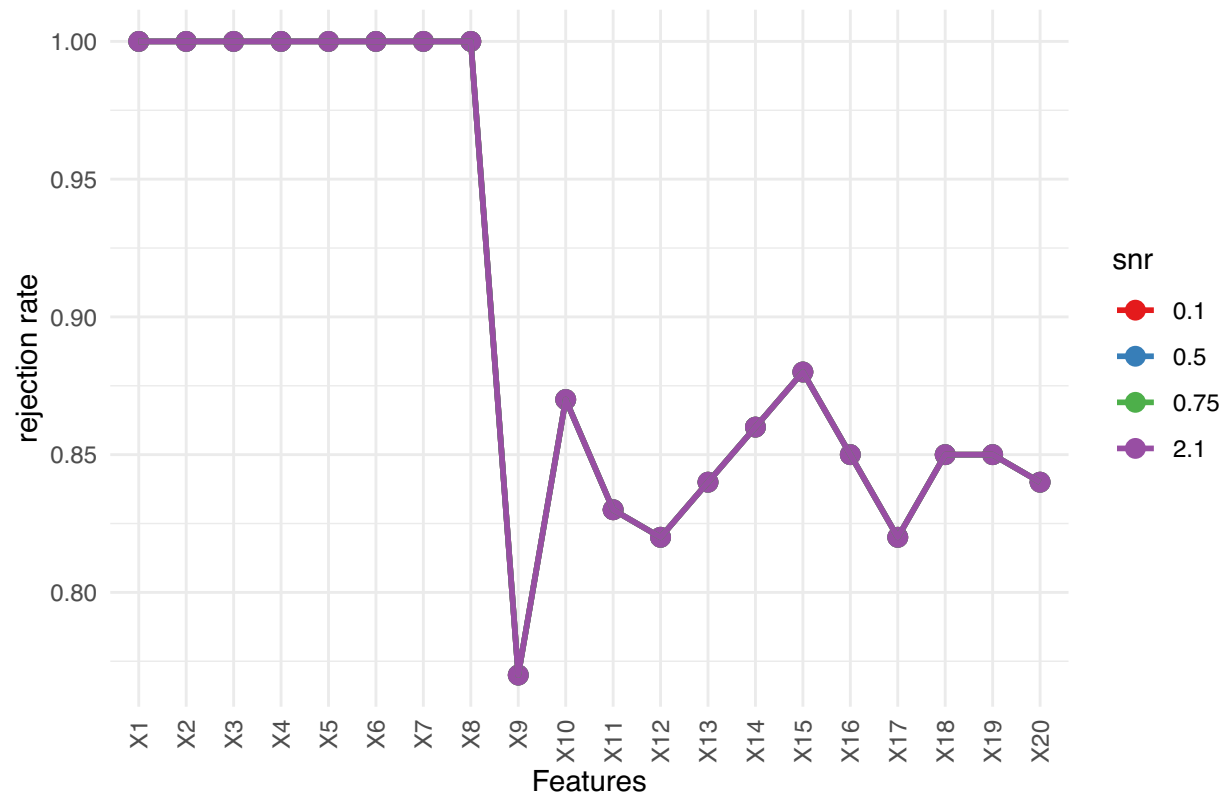
```
##      user  system elapsed
## 591.33    5.64  607.35
```

Case b):



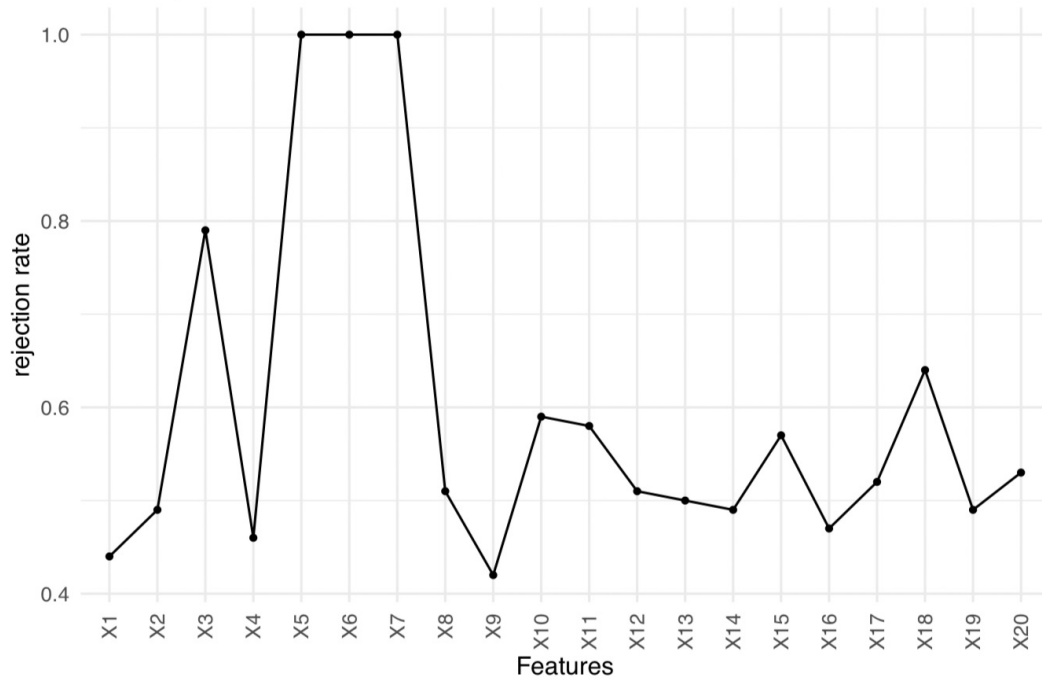
```
##      user  system elapsed
## 615.02    6.28  632.23
```

Case c):



```
##      user  system elapsed
## 371.30    3.06  380.27
```

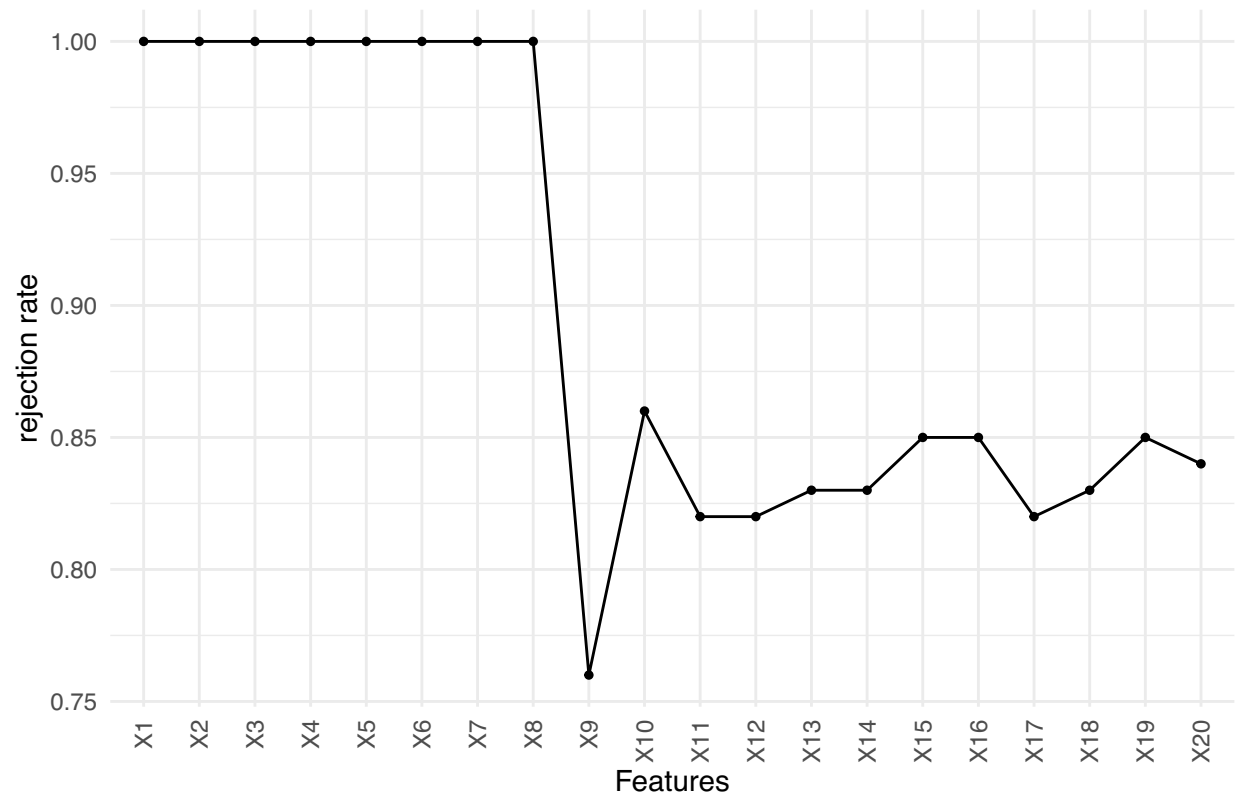
Case d):



LoGo subset selection
(100 simulations, with each 500 instances).

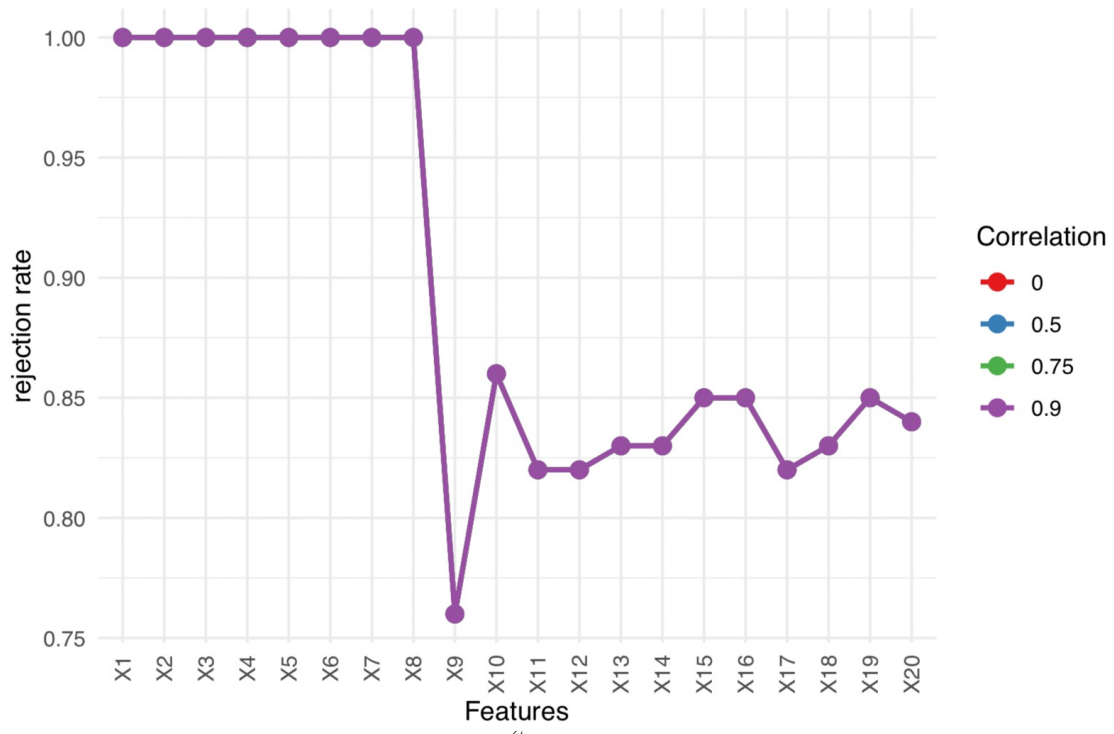
```
##      user  system elapsed
##    99.47    0.51   101.44
```

Case a):



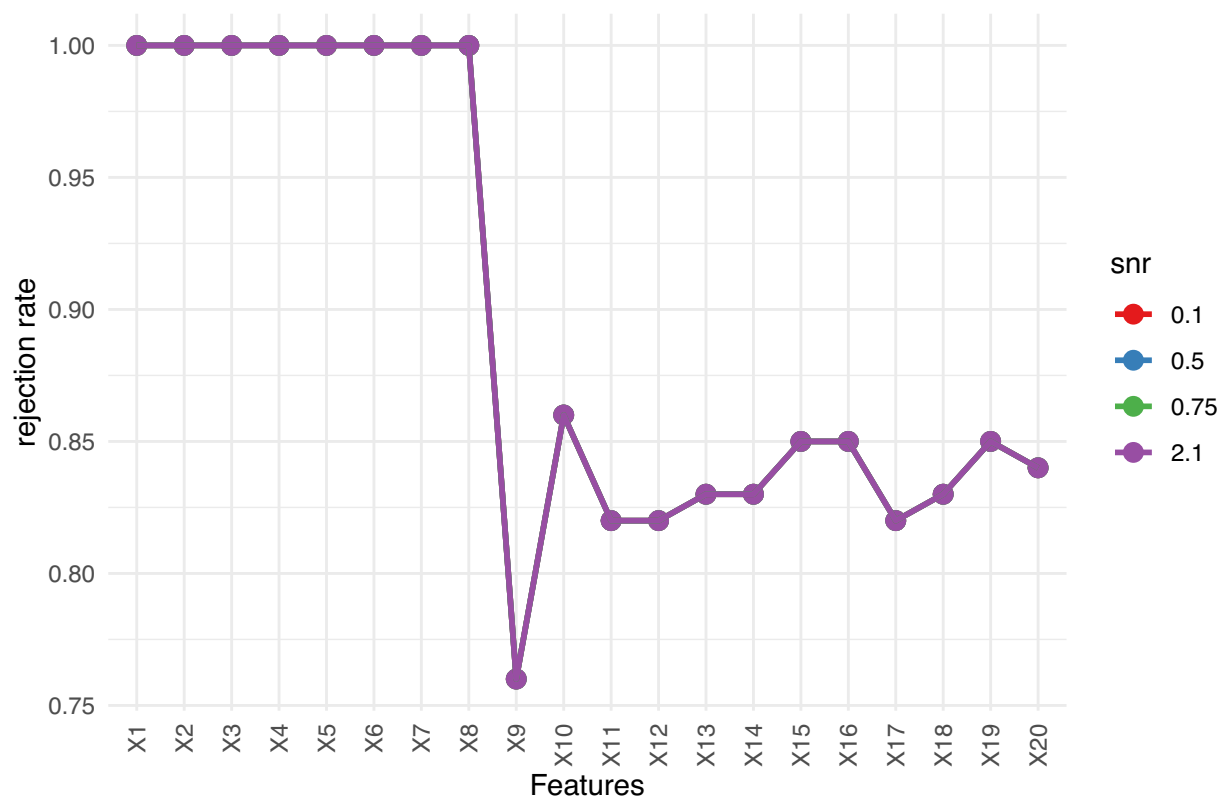
```
##      user  system elapsed
##   392.16    2.19   399.04
```

Case b):



```
##      user  system elapsed
## 340.82    1.56  345.58
```

Case c):



```
##      user  system elapsed
## 99.70    0.34  101.28
```

Case d):

