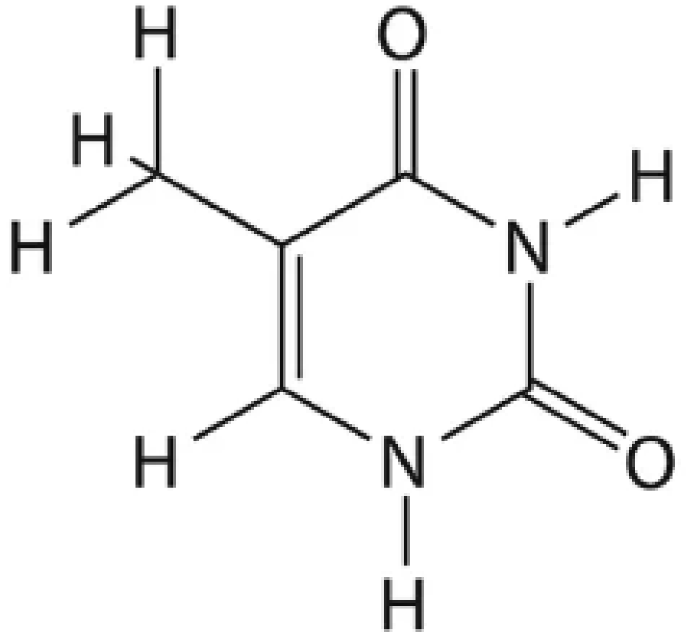
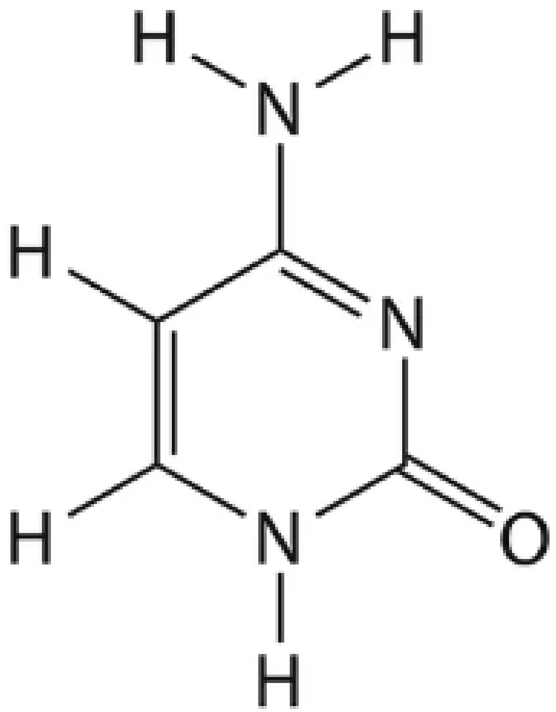
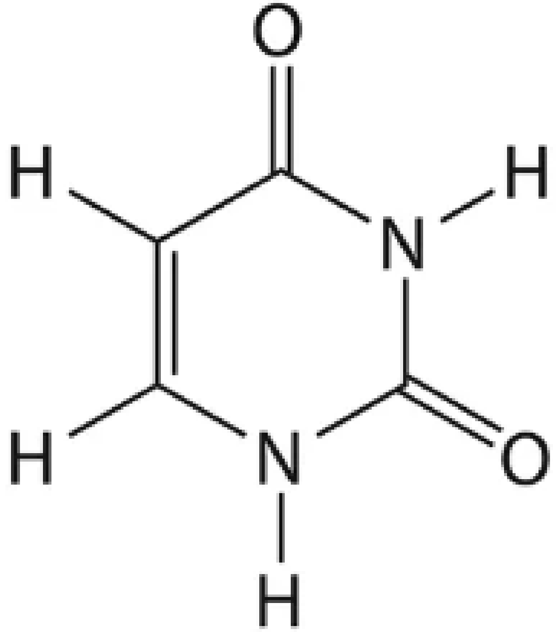




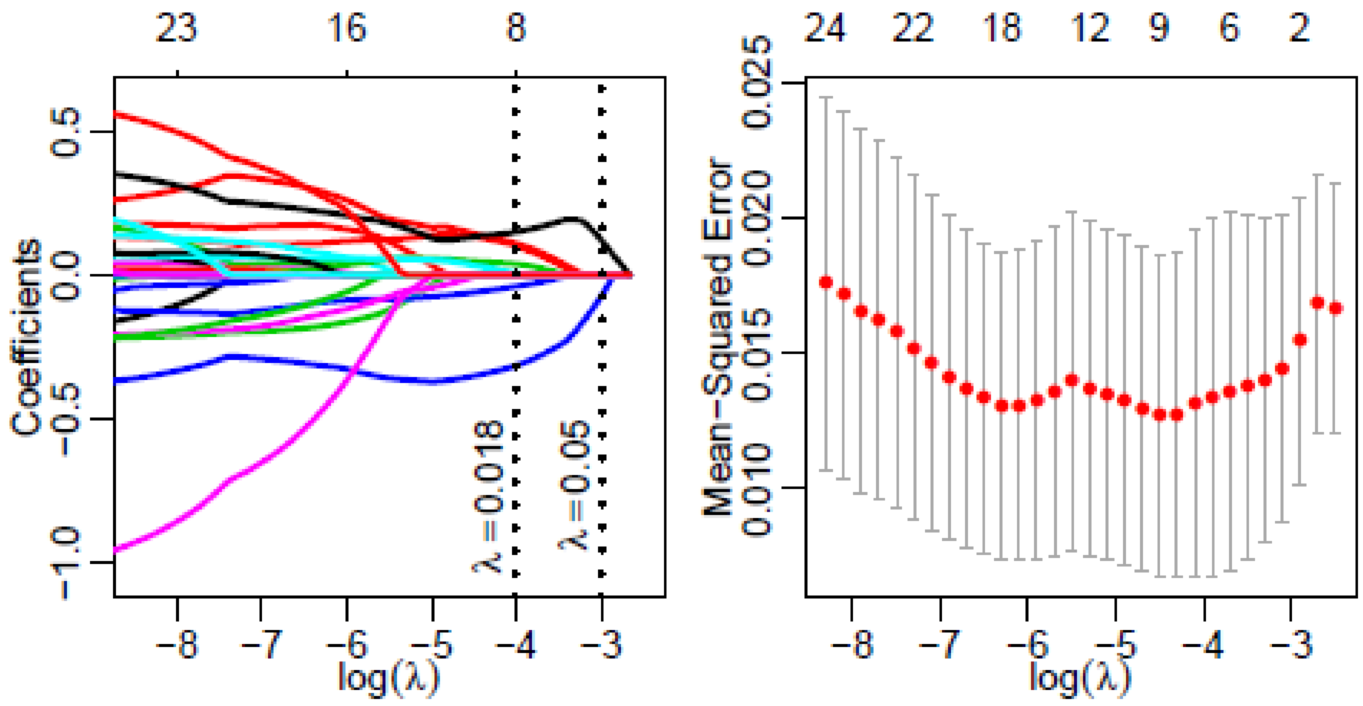


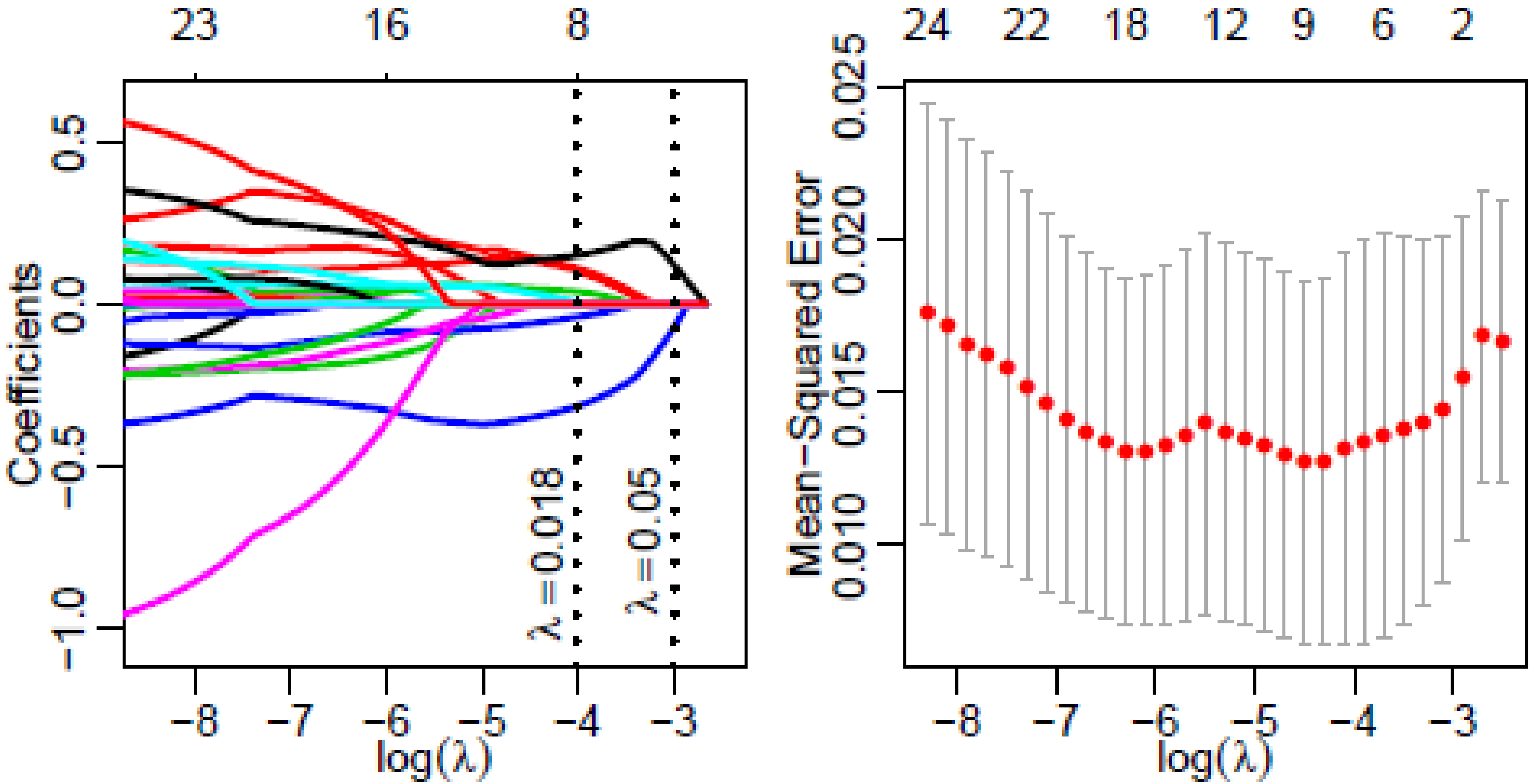
Exercise 19

Presenters: Oriane Kopp and Myroslava Volosko

Questioners: Anja Probst and Jazmin Valeriano

We again consider the pyrimidine data set presented in the lecture. It contains 74 activity measurements of the enzyme DHFR in a bacterium in the presence of different pyrimidines characterized by 26 physico-chemical properties. Those properties are quantified by the variables X1 to X26; the activity of DHFR is the response variable Y. The following figure shows the regularization path (left) and the result of a leave-one-out cross validation (right):



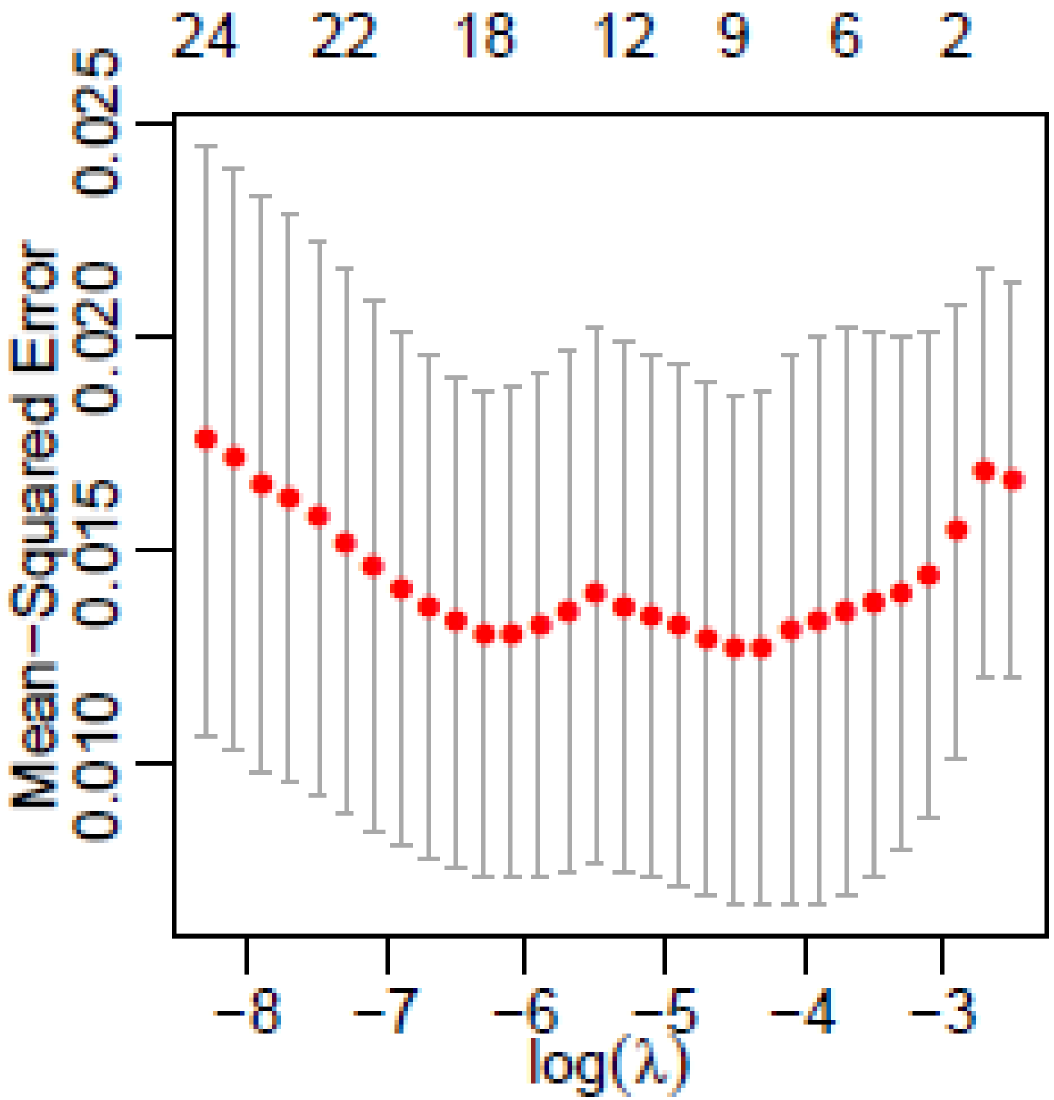

23
16
8
0.5
0.0
−0.5
−1.0
Coefficients
λ = 0.018
λ = 0.05
−8
−7
−6
−5
−4
−3
log(λ)
24
22
18
12
9
6
2
0.025
0.020
0.015
0.010
Mean−Squared Error
−8
−7
−6
−5
−4
−3
log(λ)

LASSO 🤠

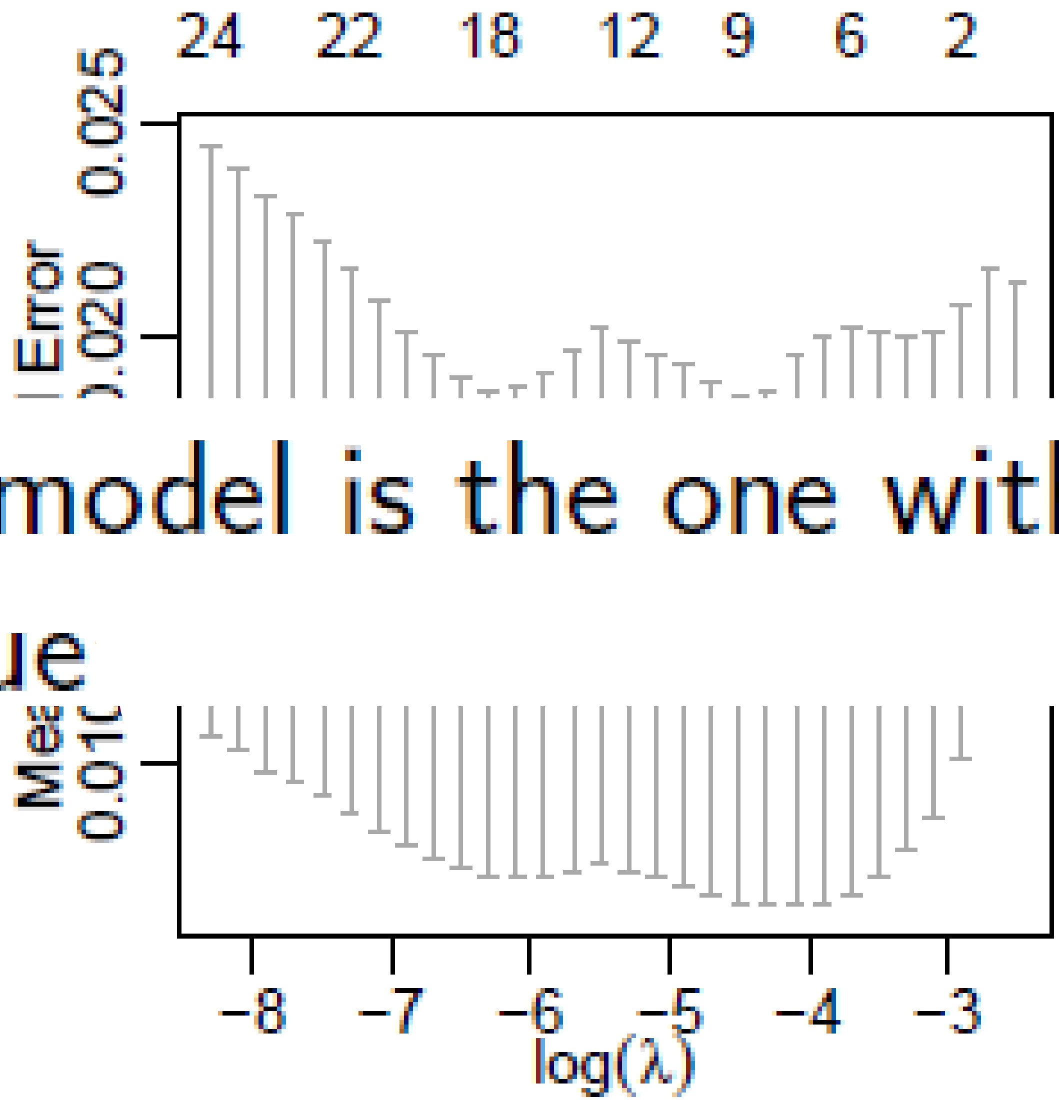
- Least Absolute Shrinkage and Selection Operator
- Find vector of coefficients β that minimizes the penalized residual sum of squares

$$\mathrm{RSS} + \lambda \sum_{j=1}^{p} |\beta_j| = \|Y - \mathbf{X}\beta\|^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

a) What is the optimal range for the regularization parameter ?



a) What is the optimal range for the regularization parameter ?



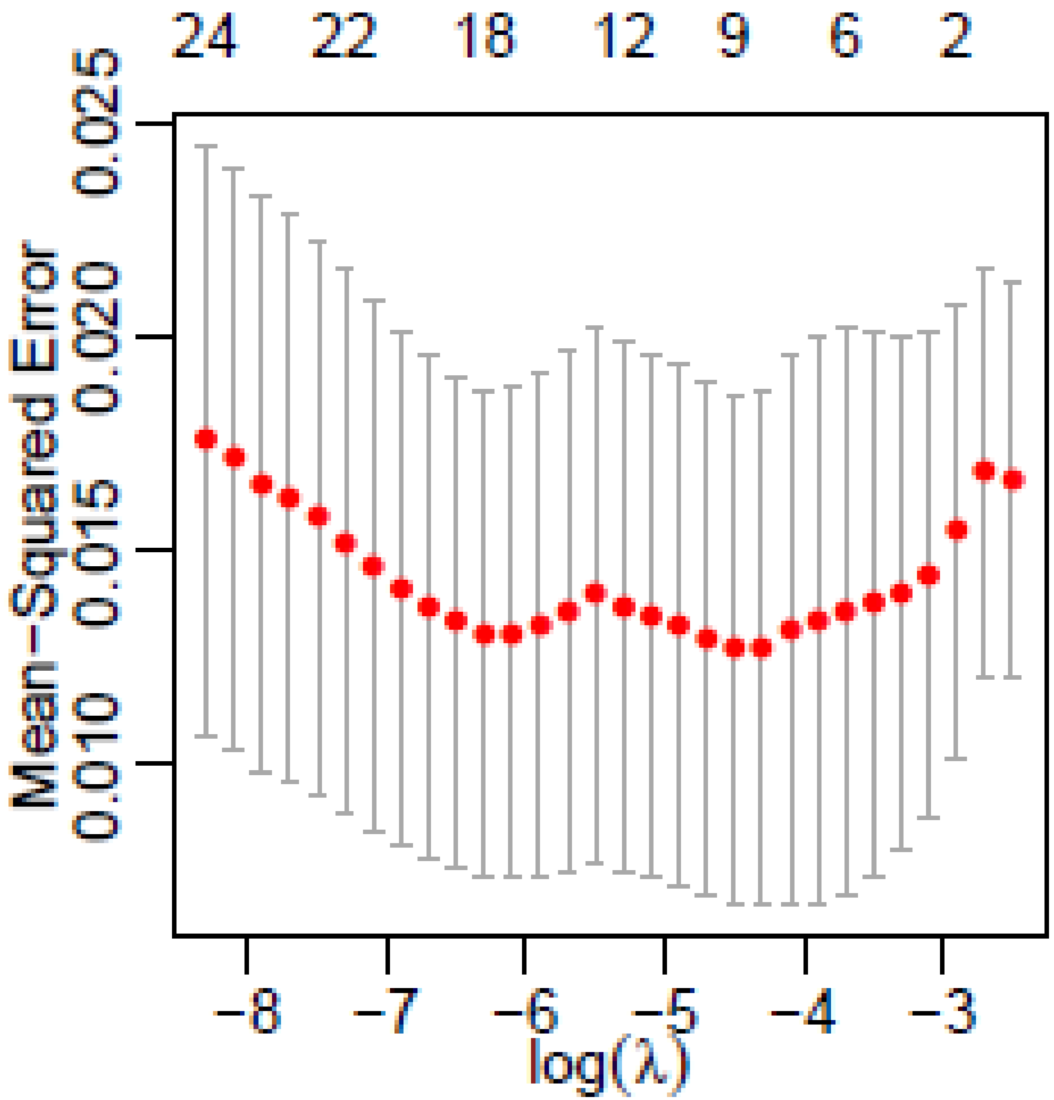
"Best" model is the one with minimum CV-value

What is CV-value?

- (Leave-one-out) Cross-validation value
- Remove one data point of the dataset and fit the model on the remaining data points → test the model on the data point you put aside → repeat

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2$$

a) What is the optimal range for the regularization parameter ?



a) What is the optimal range for the regularization parameter ?

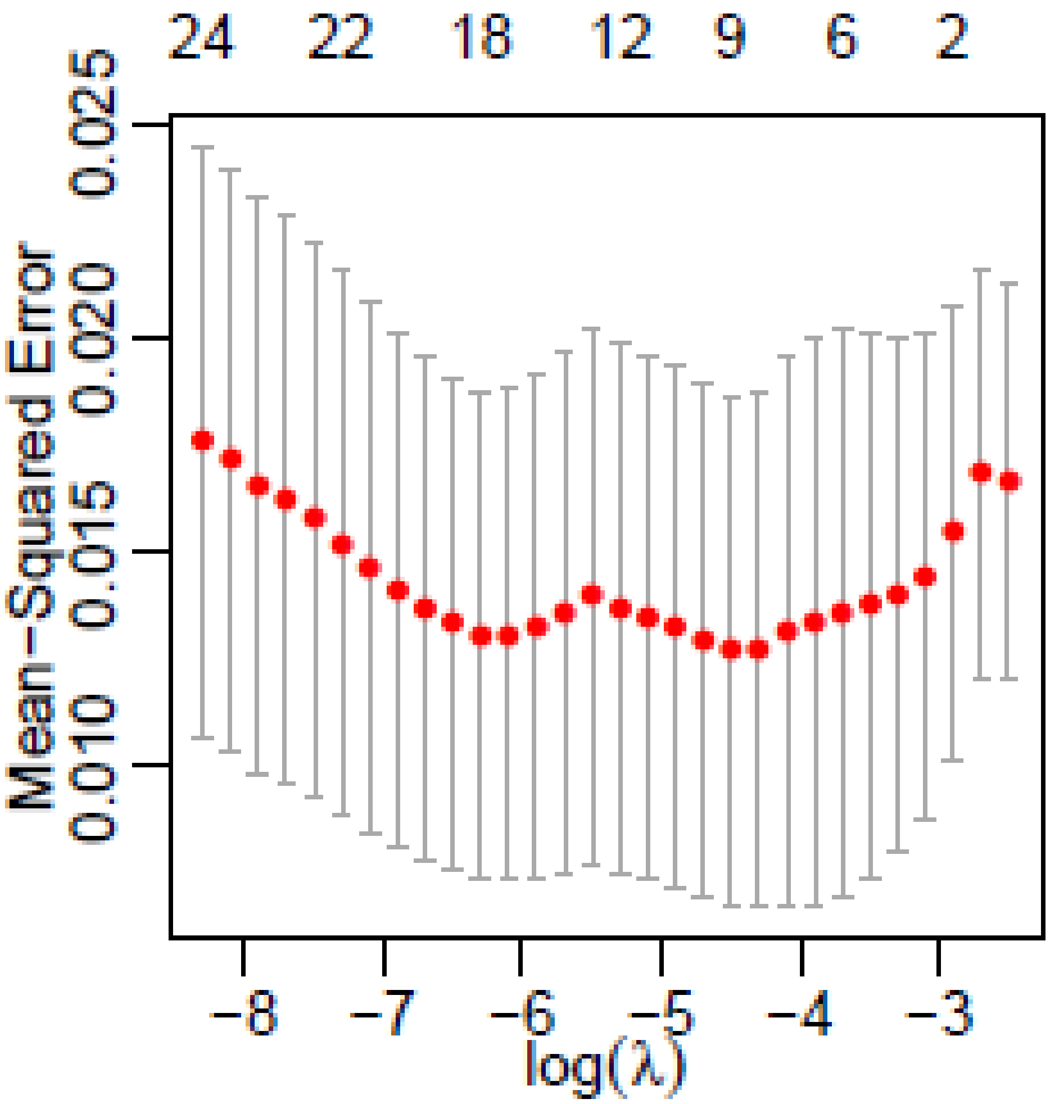
[-5, -3] → lambda should be positive

[0.0001, 0.001] → in log [-9, -7]

[0.005, 0.05] → in log [-5.2, -3]

[0.1, 1] → in log [-2, 0]

[3, 5] → in log [1, 1.6]



a) What is the optimal range for the regularization parameter ?

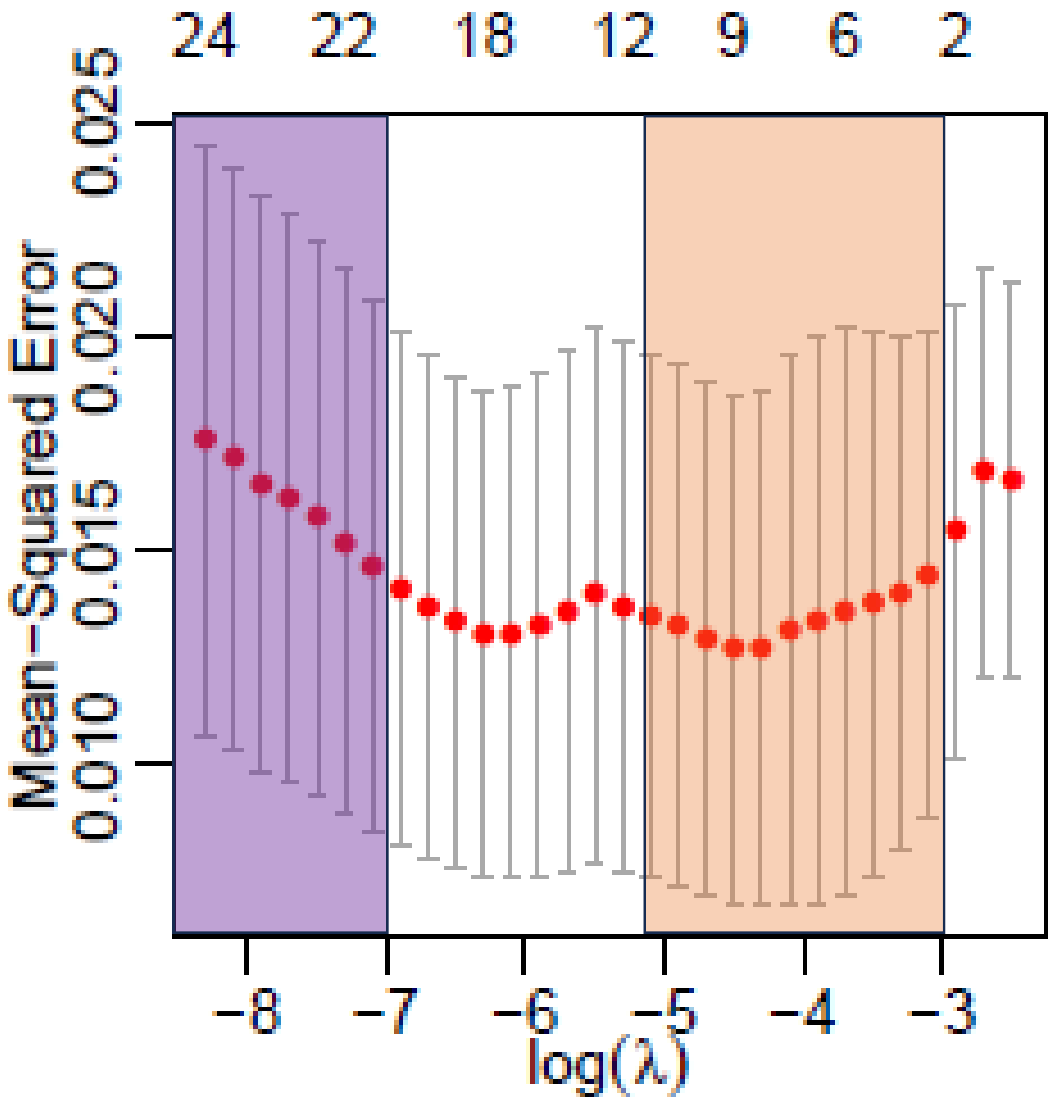
~~[-5, -3] → lambda should be positive~~

[0.0001, 0.001] → in log [-9, -7]

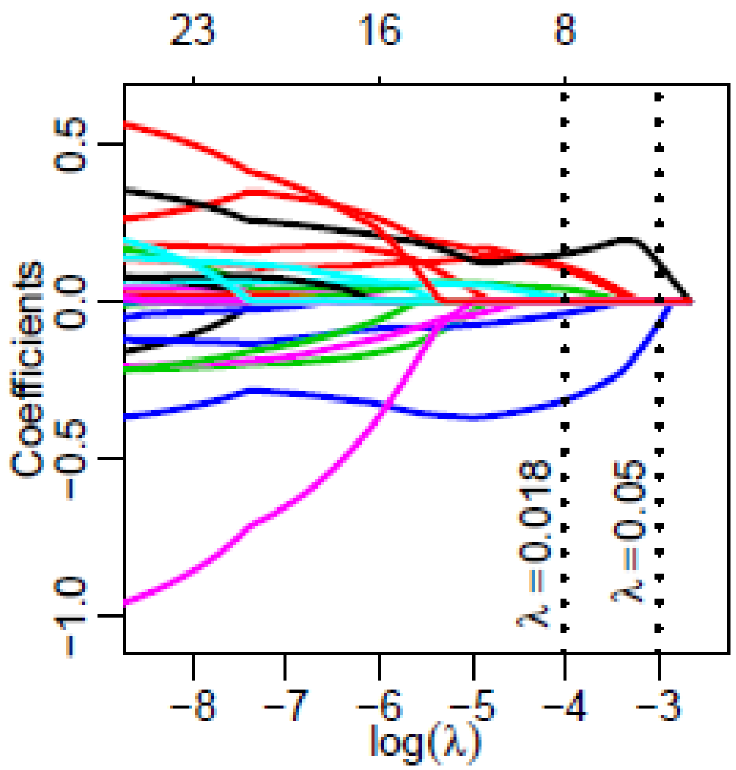
[0.005, 0.05] → in log [-5.2, -3]

~~[0.1, 1] → in log [-2, 0]~~

~~[3, 5] → in log [1, 1.6]~~



b) When choosing λ = 0.018 (left dotted line in left plot), we get the following coefficients (rounded):




```
27 x 1 sparse Matrix of class "dgCMatrix"
                    1
(Intercept)  0.600
x1           .
x2           .
x3           .
x4          -0.040
x5           .
x6           .
x7           .
x8           0.103
x9           0.001
x10          .
x11          .
x12          .
x13          .
x14          .
x15          0.039
x16          .
x17          0.008
x18          .
x19          .
x20          0.112
x21          .
x22         -0.316
x23          .
x24          .
x25          0.150
x26          .
```

b) How many variables are selected in this model? Write down the corresponding model equation.

```
27 x 1 sparse Matrix of class "dgCMatrix"
                   1
(Intercept)   0.600
x1            .
x2            .
x3            .
x4           -0.040
x5            .
x6            .
x7            .
x8            0.103
x9            0.001
x10           .
x11           .
x12           .
x13           .
x14           .
x15           0.039
x16           .
x17           0.008
x18           .
x19           .
x20           0.112
x21           .
x22          -0.316
x23           .
x24           .
x25           0.150
x26           .
```

b) How many variables are selected in this model? Write down the corresponding model equation.

```
27 x 1 sparse Matrix of class "dgCMatrix"
                      1
(Intercept)   0.600
x1            .
x2            .
x3            .
x4           -0.040
x5            .
x6            .
x7            .
x8            0.103
x9            0.001
x10           .
x11           .
x12           .
x13           .
x14           .
x15           0.039
x16           .
x17           0.008
x18           .
x19           .
x20           0.112
x21           .
x22          -0.316
x23           .
x24           .
x25           0.150
x26           .
```

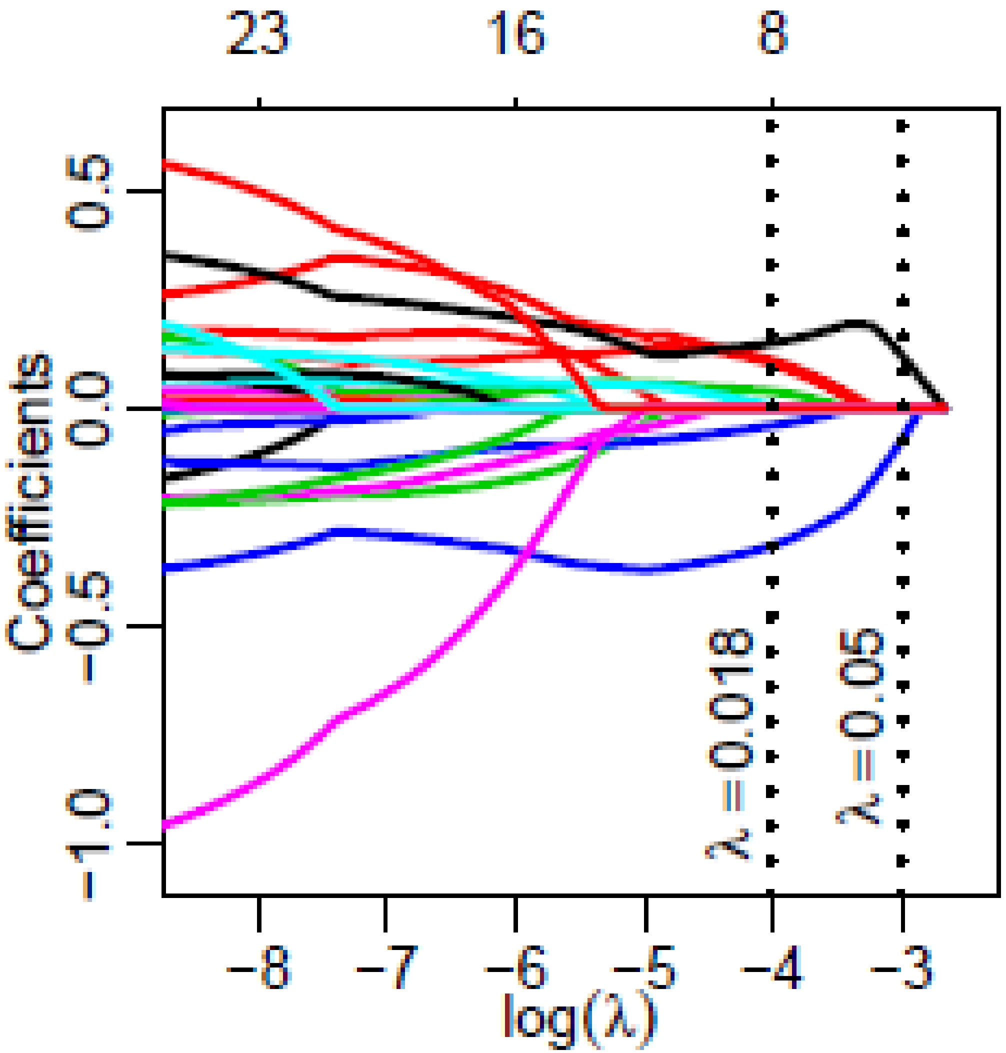
Only keep the variables that have nonzero coefficients

b) How many variables are selected in this model? Write down the corresponding model equation.

```
27 x 1 sparse Matrix of class "dgCMatrix"
                  1
(Intercept)   0.600
x1             .
x2             .
x3             .
x4           -0.040
x5             .
x6             .
x7             .
x8            0.103
x9            0.001
x10            .
x11            .
x12            .
x13            .
x14            .
x15           0.039
x16            .
x17           0.008
x18            .
x19            .
x20           0.112
x21            .
x22          -0.316
x23            .
x24            .
x25           0.150
x26            .
```

$$Y = 0.600 - 0.040x_4 + 0.103x_8 + 0.001x_9 + 0.039x_{15} + 0.008x_{17} + 0.112x_{20} - 0.316x_{22} + 0.150x_{25} + E,\ E \sim N(0, \sigma^2)$$

c) Assume now that we choose λ= 0.05 (right dotted line in left plot). Which variables remain in the model in this case?




```
27 x 1 sparse Matrix of class "dgCMatrix"
                 1
(Intercept)  0.600
x1           .
x2           .
x3           .
x4          -0.040
x5           .
x6           .
x7           .
x8           0.103
x9           0.001
x10          .
x11          .
x12          .
x13          .
x14          .
x15          0.039
x16          .
x17          0.008
x18          .
x19          .
x20          0.112
x21          .
x22         -0.316
x23          .
x24          .
x25          0.150
x26          .
```

What do we know? 💡

- Values with coefficient 0 with a bigger lambda won't suddenly reappear with a smaller lambda, so we know from b) that only x4, x8, x9, x15, x17, x20, x22 or x25 can still be in the model

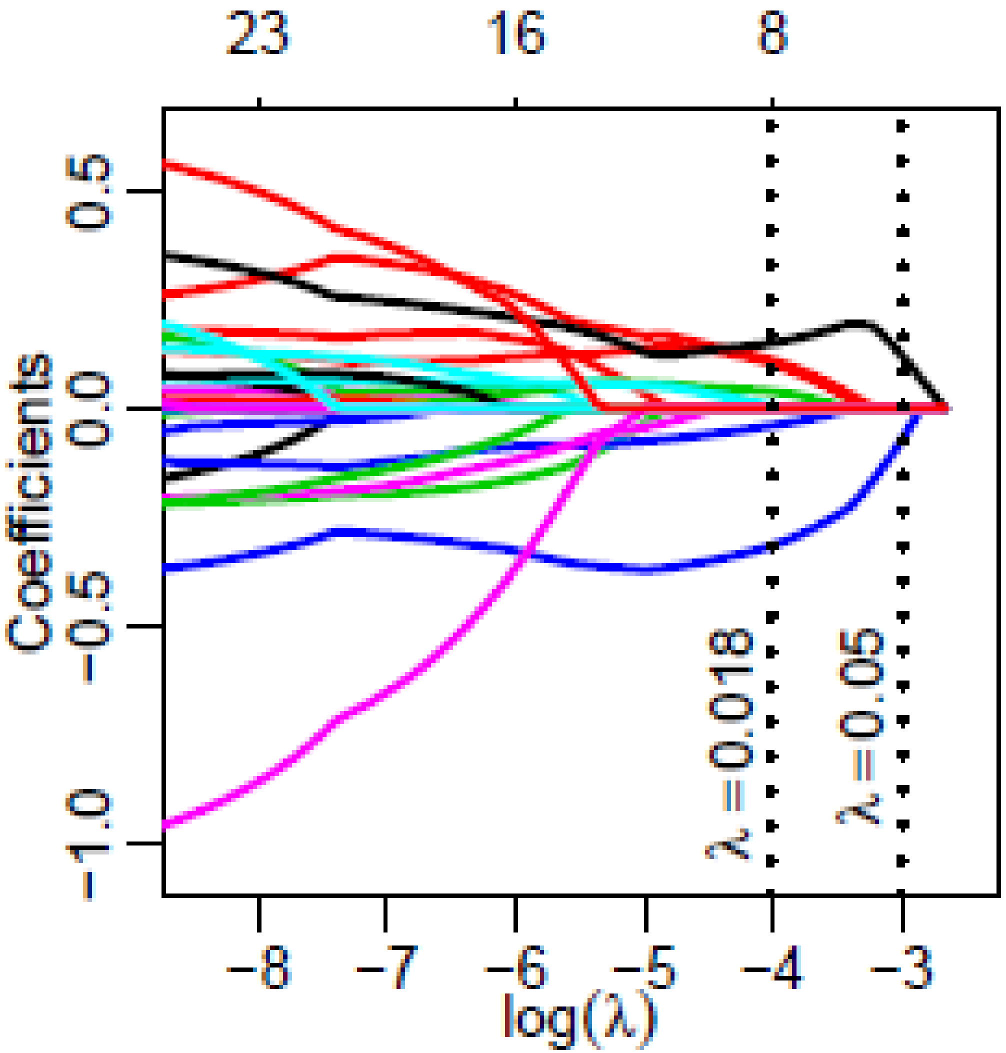
- X4, X8, X22
- X4, X20
- X14, X22, X25
- X15, X17, X22
- X22, X25

What do we know? 💡

- Values with coefficient 0 with a bigger lambda won't suddenly reappear with a smaller lambda, so we know from b) that only x4, x8, x9, x15, x17, x20, x22 or x25 can still be in the model

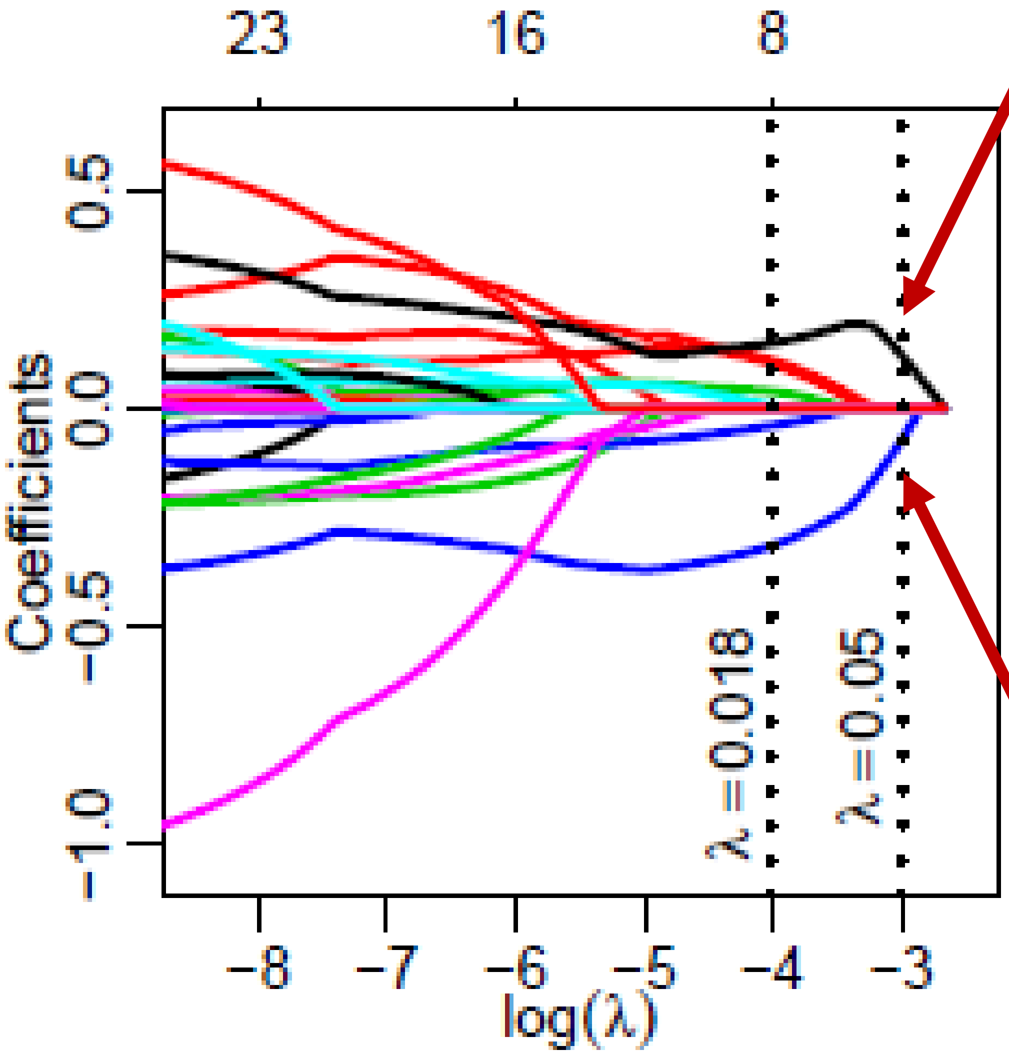
- X4, X8, X22
- X4, X20
- ~~X14, X22, X25~~
- X15, X17, X22
- X22, X25

c) Assume now that we choose λ= 0.05 (right dotted line in left plot). Which variables remain in the model in this case?




```
27 x 1 sparse Matrix of class "dgCMatrix"
                   1
(Intercept)    0.600
x1                 .
x2                 .
x3                 .
x4            -0.040
x5                 .
x6                 .
x7                 .
x8             0.103
x9             0.001
x10                .
x11                .
x12                .
x13                .
x14                .
x15            0.039
x16                .
x17            0.008
x18                .
x19                .
x20            0.112
x21                .
x22           -0.316
x23                .
x24                .
x25            0.150
x26                .
```

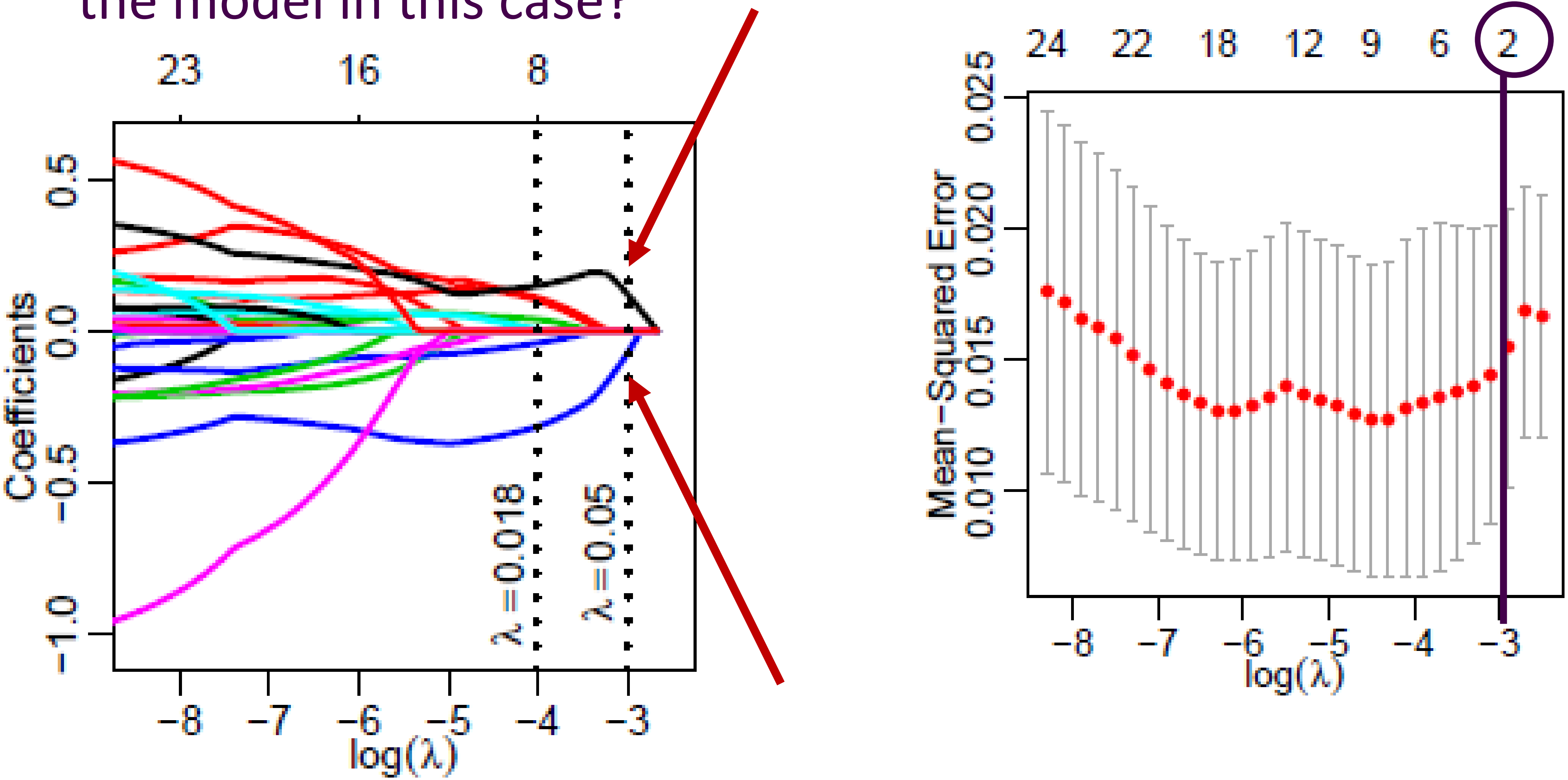
c) Assume now that we choose λ= 0.05 (right dotted line in left plot). Which variables remain in the model in this case?



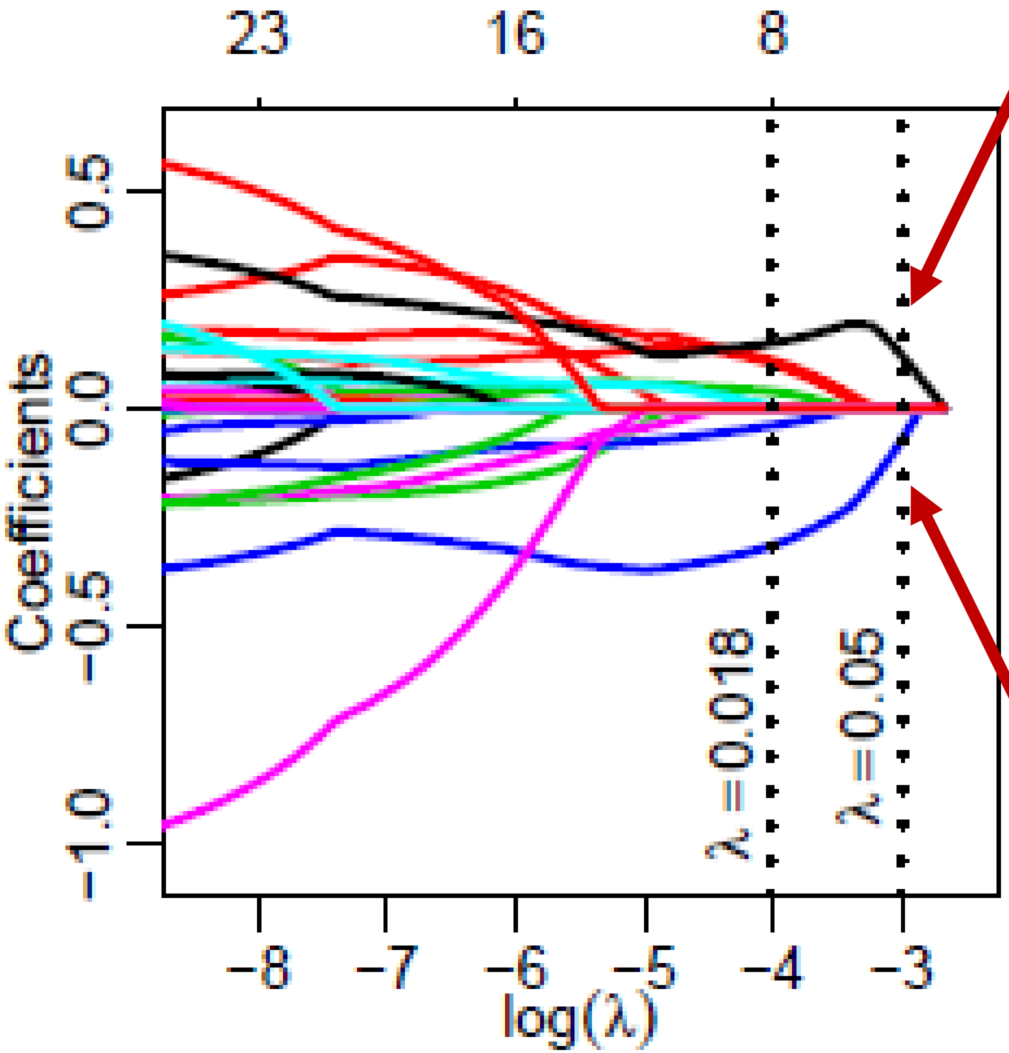

```
27 x 1 sparse Matrix of class "dgCMatrix"
                  1
(Intercept)  0.600
x1           .
x2           .
x3           .
x4          -0.040
x5           .
x6           .
x7           .
x8           0.103
x9           0.001
x10          .
x11          .
x12          .
x13          .
x14          .
x15          0.039
x16          .
x17          0.008
x18          .
x19          .
x20          0.112
x21          .
x22         -0.316
x23          .
x24          .
x25          0.150
x26          .
```

X4, X8, X22
X4, X20
~~X14, X22, X25~~
X15, X17, X22
X22, X25

c) Assume now that we choose λ= 0.05 (right dotted line in left plot). Which variables remain in the model in this case?



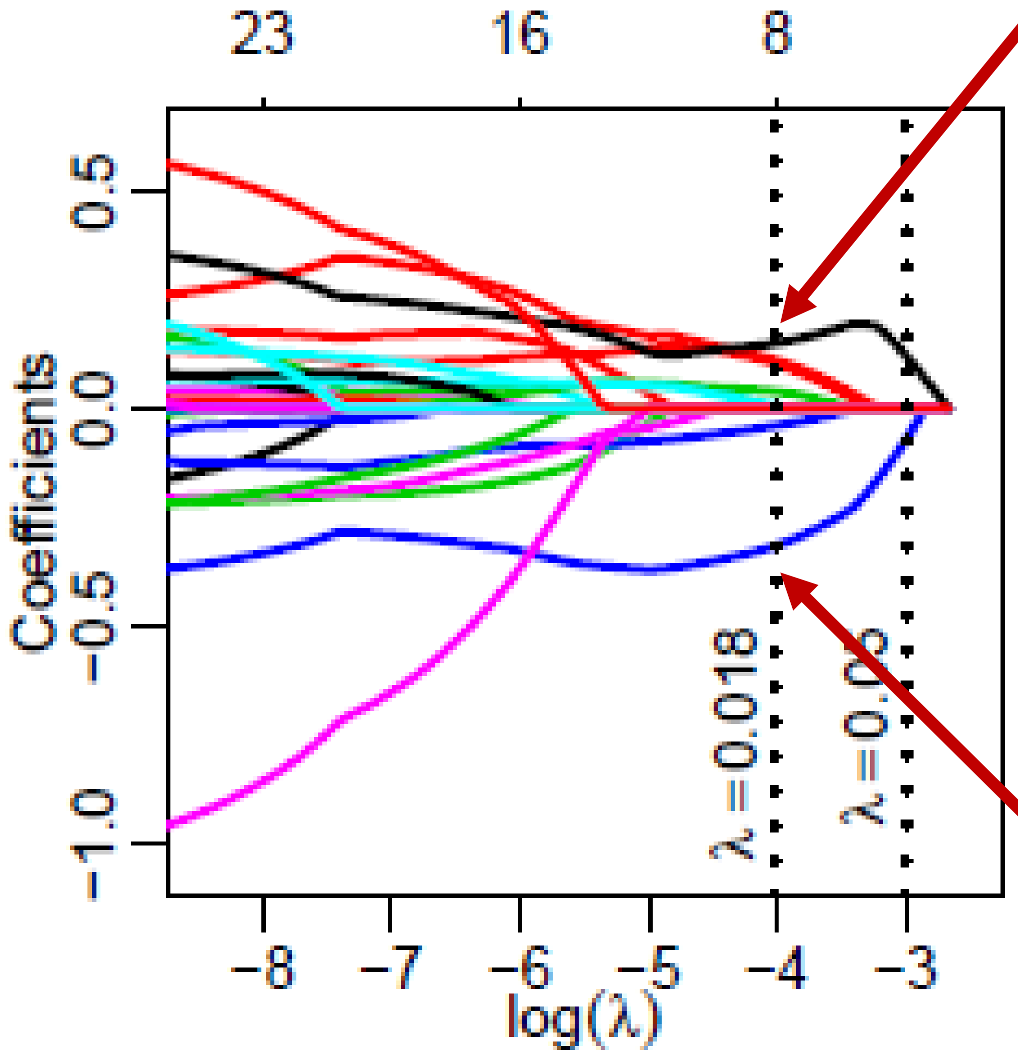
c) Assume now that we choose λ= 0.05 (right dotted line in left plot). Which variables remain in the model in this case?




```
27 x 1 sparse Matrix of class "dgCMatrix"
                  1
(Intercept)  0.600
x1            .
x2            .
x3            .
x4          -0.040
x5            .
x6            .
x7            .
x8           0.103
x9           0.001
x10           .
x11           .
x12           .
x13           .
x14           .
x15          0.039
x16           .
x17          0.008
x18           .
x19           .
x20          0.112
x21           .
x22         -0.316
x23           .
x24           .
x25          0.150
x26           .
```

~~X4, X8, X22~~
X4, X20
~~X14, X22, X25~~
~~X15, X17, X22~~
X22, X25

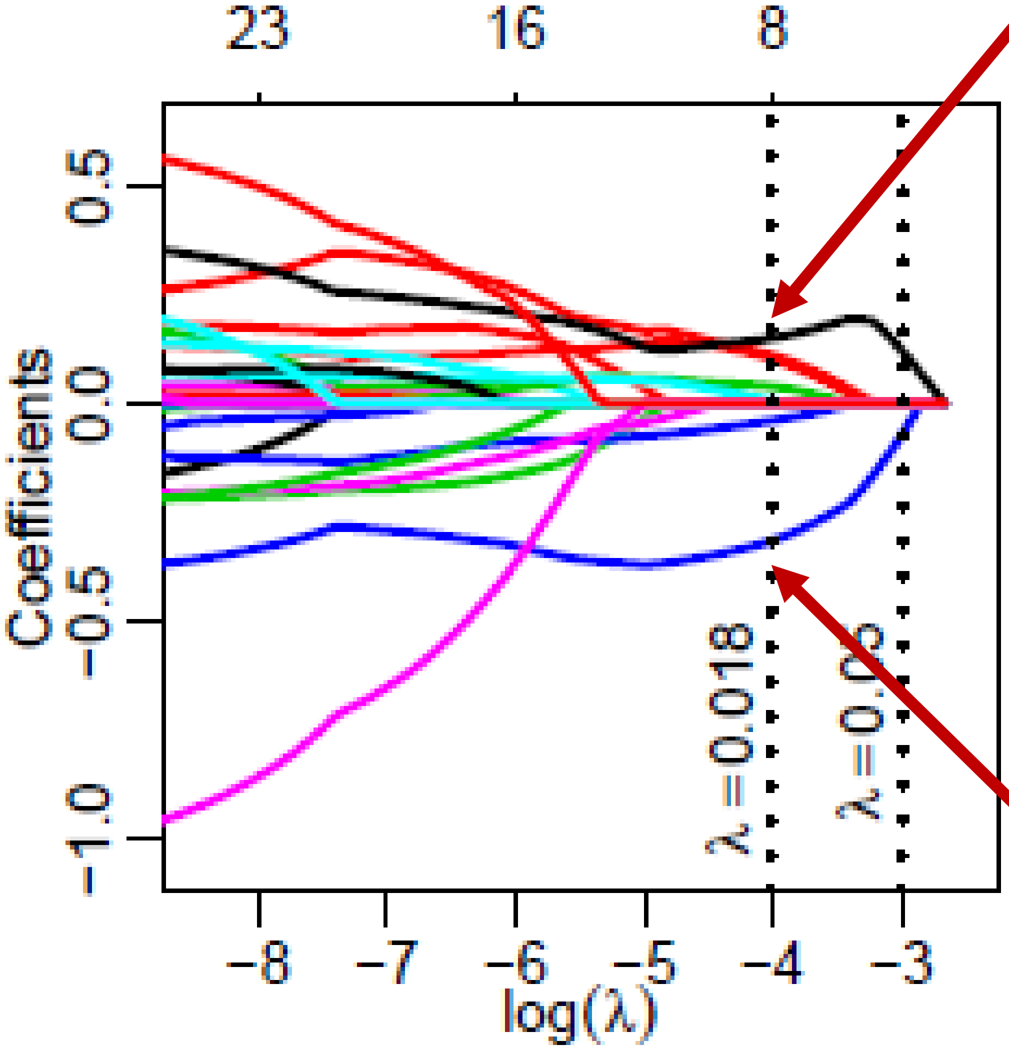
c) Assume now that we choose λ= 0.05 (right dotted line in left plot). Which variables remain in the model in this case?




```
27 x 1 sparse Matrix of class "dgCMatrix"
                 1
(Intercept)  0.600
x1           .
x2           .
x3           .
x4          -0.040
x5           .
x6           .
x7           .
x8           0.103
x9           0.001
x10          .
x11          .
x12          .
x13          .
x14          .
x15          0.039
x16          .
x17          0.008
x18          .
x19          .
x20          0.112
x21          .
x22         -0.316
x23          .
x24          .
x25          0.150
x26          .
```

Only the variables with the biggest and smallest coefficient in model from b) remains

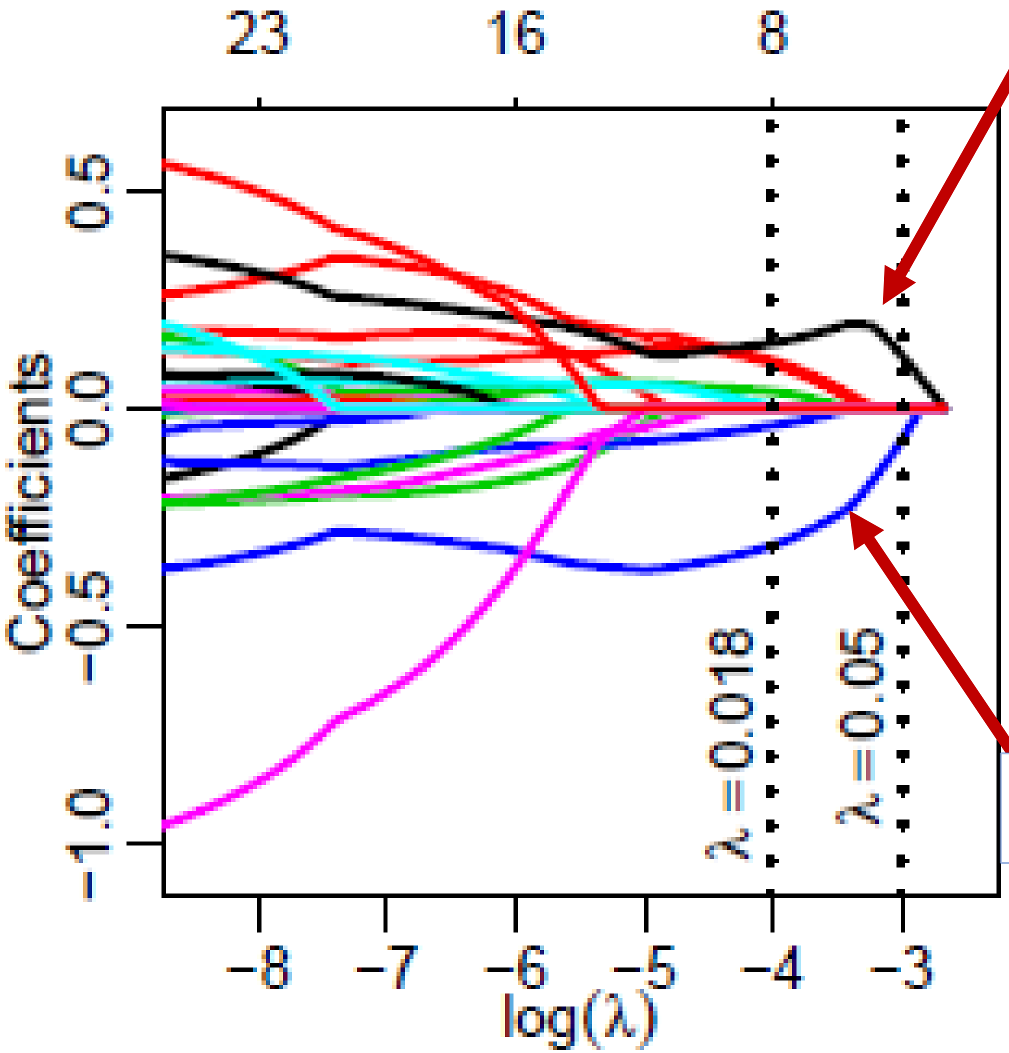
c) Assume now that we choose λ= 0.05 (right dotted line in left plot). Which variables remain in the model in this case?




```
27 x 1 sparse Matrix of class "dgCMatrix"
                   1
(Intercept)   0.600
x1            .
x2            .
x3            .
x4           -0.040
x5            .
x6            .
x7            .
x8            0.103
x9            0.001
x10           .
x11           .
x12           .
x13           .
x14           .
x15           0.039
x16           .
x17           0.008
x18           .
x19           .
x20           0.112
x21           .
x22          -0.316
x23           .
x24           .
x25           0.150
x26           .
```

~~X4, X8, X22~~
~~X4, X20~~
~~X14, X22, X25~~
~~X15, X17, X22~~
X22, X25

c) Assume now that we choose λ= 0.05 (right dotted line in left plot). Which variables remain in the model in this case?



27 x 1 sparse Matrix of class "dgCMatrix"

	1		
(Intercept)	0.600	x13	.
x1	.	x14	.
x2	.	x15	0.039
x3	.	x16	.
x4	-0.040	x17	0.008
x5	.	x18	.
x6	.	x19	.
x7	.	x20	0.112
x8	0.103	x21	.
x9	0.001	x22	-0.316
x10	.	x23	.
x11	.	x24	.
x12	.	x25	0.150
		x26	.

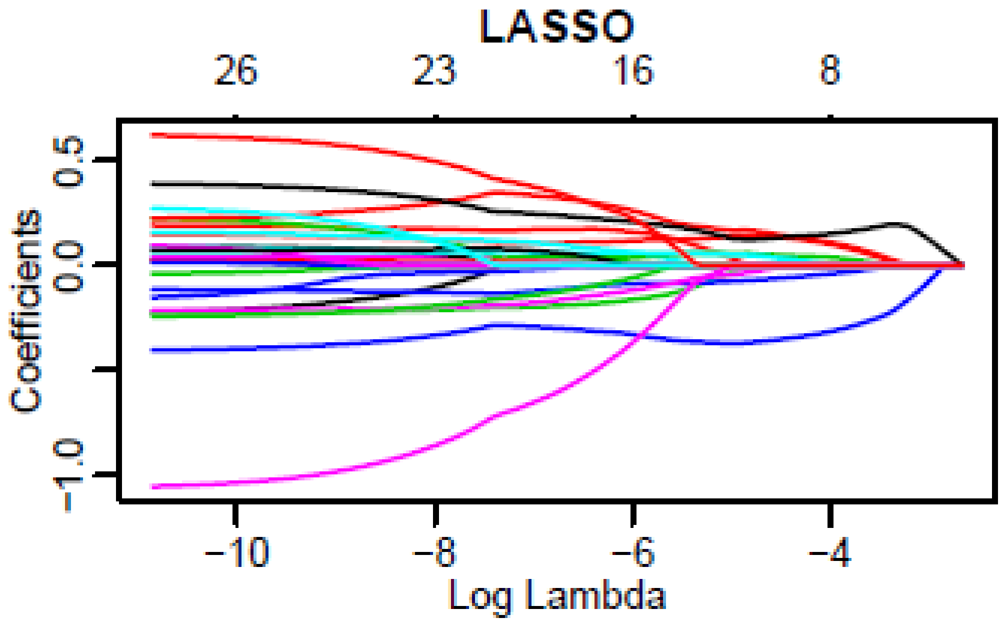
$$Y = \beta_0 + \beta_1 x25 + \beta_2 x22 + E, \;\; E \sim N(0, \sigma^2)$$

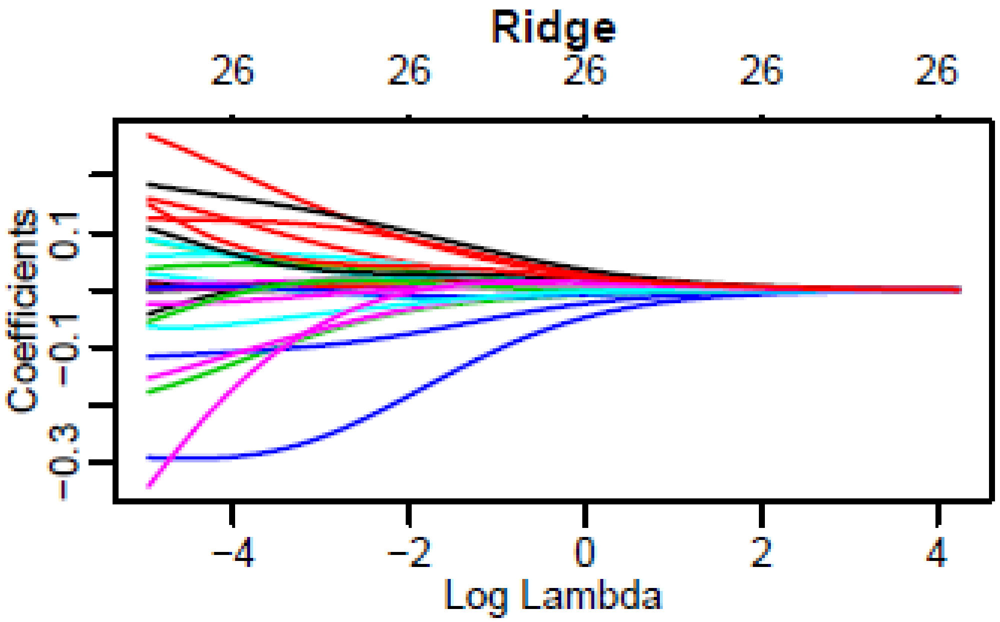
Conclusion 😳

- In this exercise, the point is not to find the correct model but to "play" with LASSO
- LASSO is great for viewing the dynamic of the coefficients of the variable
- Hard to tell the exact values for regularization parameter and path just by looking….

Questions? 🙂

From the lecture (slide 161)





K-fold vs Leave-one-out cross validation

K-fold

- Split the dataset in k subsets
- Remove one subset and fit the model on the remaining subsets
- Test the model on the set you put aside
- Repeat for each subset

Leave-one-out

- Remove one data point (from response vector and model matrix)
- Fit the model on the remaining data points
- Test the model on the data point you removed
- Repeat for each data point

K-fold vs Leave-one-out cross validation

K-fold

- Split the dataset in k subsets
- Remove one subset and fit the model on the remaining subsets
- Test the model on the set you put aside
- Repeat for each subset

→ Better for big datasets

Leave-one-out

- Remove one data point (from response vector and model matrix)
- Fit the model on the remaining data points
- Test the model on the data point you removed
- Repeat for each data point

Why does the AIC approach not work for ridge and LASSO regression?

- Ridge regression does not perform variable selection.
- LASSO is nowadays often used for high-dimensional data ($n = p$, or even $p > n$): e.g., with genetic data. AIC only works for $n >> p$

→ That's why we use cross-validation