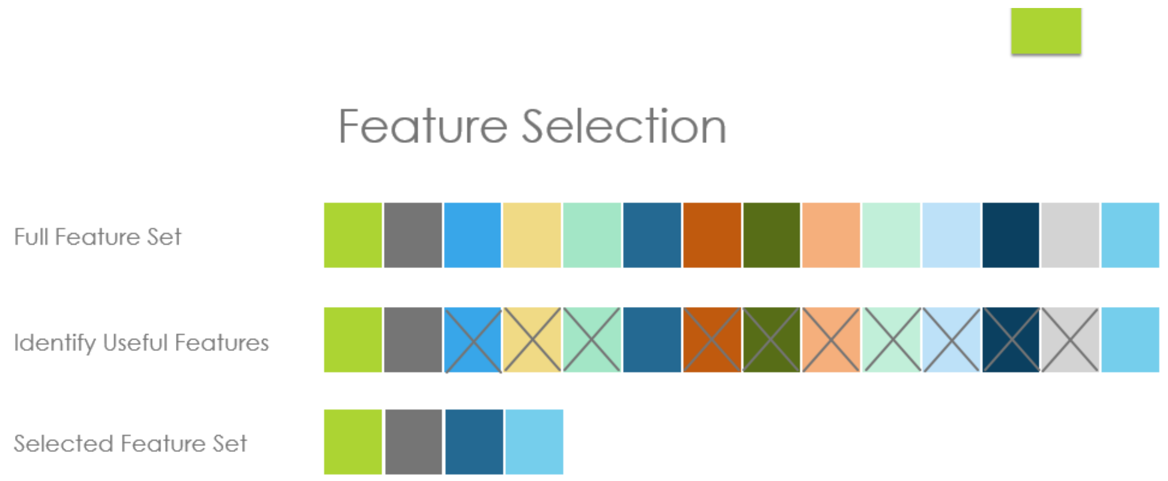
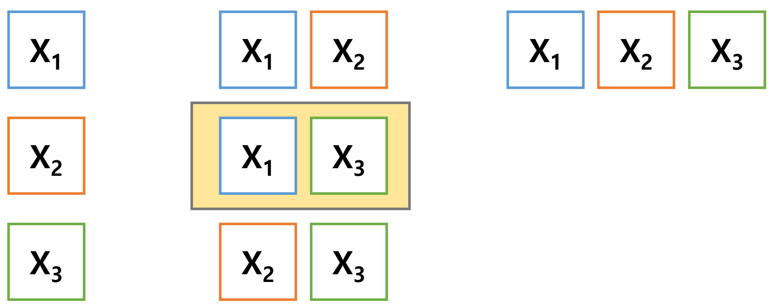


Dimensionality Reduction: Supervised Variable Selection

Pilsung Kang
School of Industrial Management Engineering
Korea University

Exhaustive Search

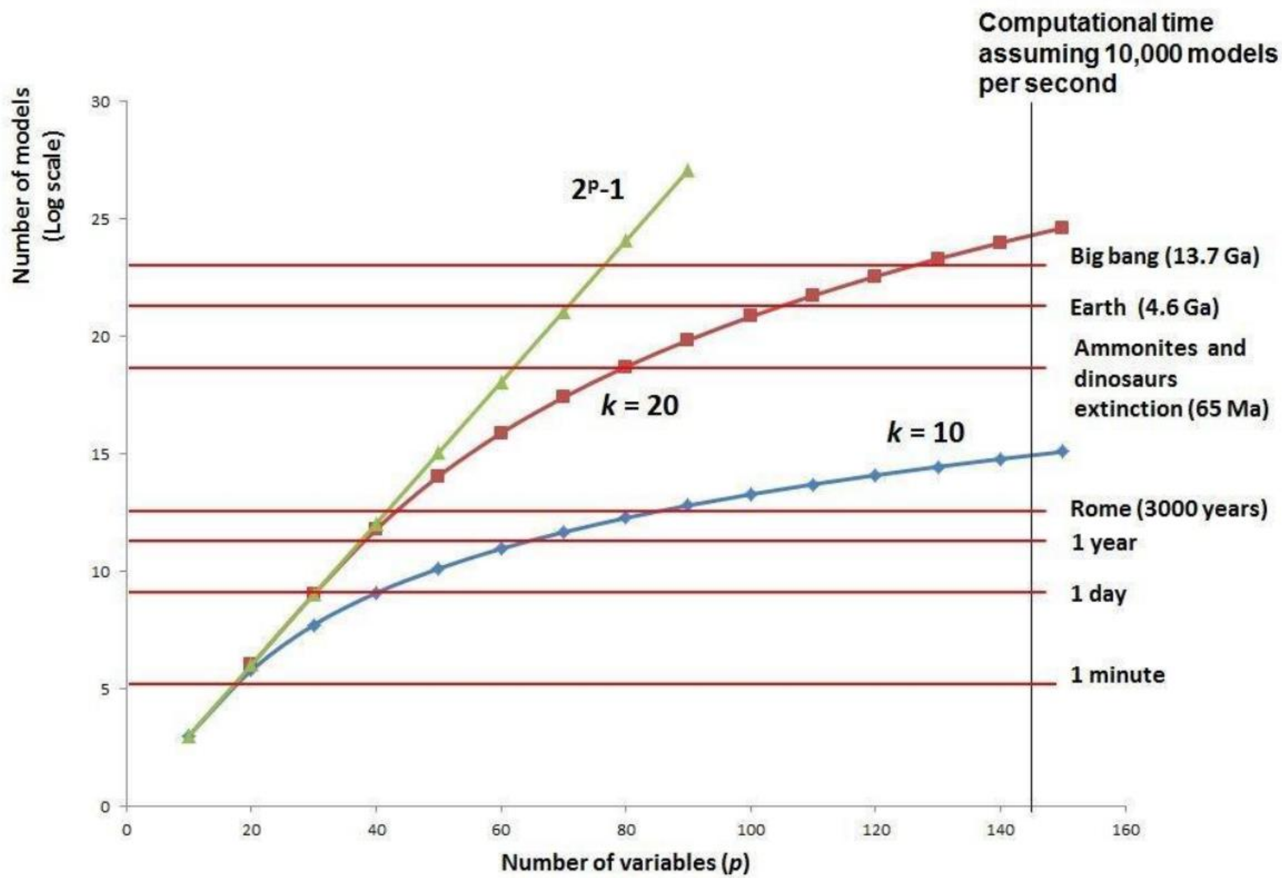
- Exhaustive search
  - ✓ Search all possible combinations
    - Ex) 3 variables X_1 X_2 X_3
    - A total of 7 possible subsets are tested



  - ✓ Performance criteria for variable selection
    - Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC), Adjusted R^2, Mallow's C_p, etc.

Exhaustive Search

- Exhaustive search
  - ✓ Assume that we have a computer that can evaluate 10,000 models/second



Forward Selection

- Forward Selection example
  - ✓ Forward Selection in the multiple linear regression
  - ✓ 8 input variables

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1, \quad R^2_{adj} = 0.48$$

$$\boxed{\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2, \quad R^2_{adj} = 0.56}$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_3 x_3, \quad R^2_{adj} = 0.51$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_4 x_4, \quad R^2_{adj} = 0.50$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_5 x_5, \quad R^2_{adj} = 0.38$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_6 x_6, \quad R^2_{adj} = 0.32$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_7 x_7, \quad R^2_{adj} = 0.50$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_8 x_8, \quad R^2_{adj} = 0.19$$

Forward Selection

- Forward Selection example
  - ✓ Forward Selection in the multiple linear regression
  - ✓ 8 input variables

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_1 x_1, \quad R^2_{adj} = 0.60$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3, \quad R^2_{adj} = 0.64$$

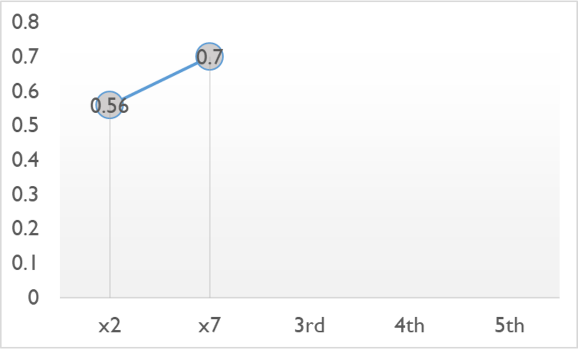
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_4 x_4, \quad R^2_{adj} = 0.58$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_5 x_5, \quad R^2_{adj} = 0.61$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_6 x_6, \quad R^2_{adj} = 0.57$$

$$\boxed{\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_7 x_7, \quad R^2_{adj} = 0.70}$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_8 x_8, \quad R^2_{adj} = 0.56$$



Forward Selection

- Forward Selection example
  - ✓ Forward Selection in the multiple linear regression
  - ✓ 8 input variables

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_7 x_7 + \hat{\beta}_1 x_1, \quad R^2_{adj} = 0.71$$

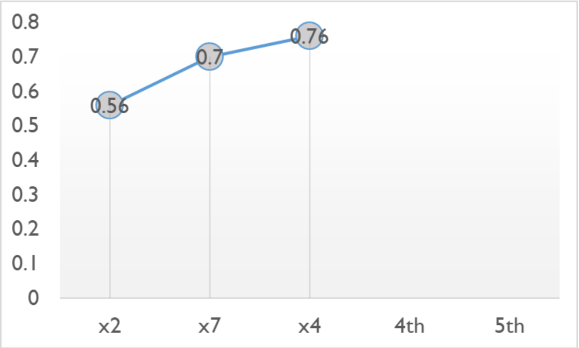
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_7 x_7 + \hat{\beta}_3 x_3, \quad R^2_{adj} = 0.72$$

$$\boxed{\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_7 x_7 + \hat{\beta}_4 x_4, \quad R^2_{adj} = 0.76}$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_7 x_7 + \hat{\beta}_5 x_5, \quad R^2_{adj} = 0.73$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_7 x_7 + \hat{\beta}_6 x_6, \quad R^2_{adj} = 0.69$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_7 x_7 + \hat{\beta}_8 x_8, \quad R^2_{adj} = 0.70$$



Forward Selection

- Forward Selection example
  - ✓ Forward Selection in the multiple linear regression
  - ✓ 8 input variables

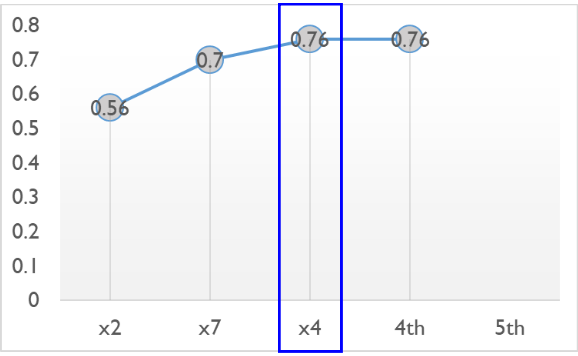
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_7 x_7 + \hat{\beta}_4 x_4 + \hat{\beta}_1 x_1, \quad R^2_{adj} = 0.76$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_7 x_7 + \hat{\beta}_4 x_4 + \hat{\beta}_3 x_3, \quad R^2_{adj} = 0.76$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_7 x_7 + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5, \quad R^2_{adj} = 0.75$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_7 x_7 + \hat{\beta}_4 x_4 + \hat{\beta}_6 x_6, \quad R^2_{adj} = 0.76$$

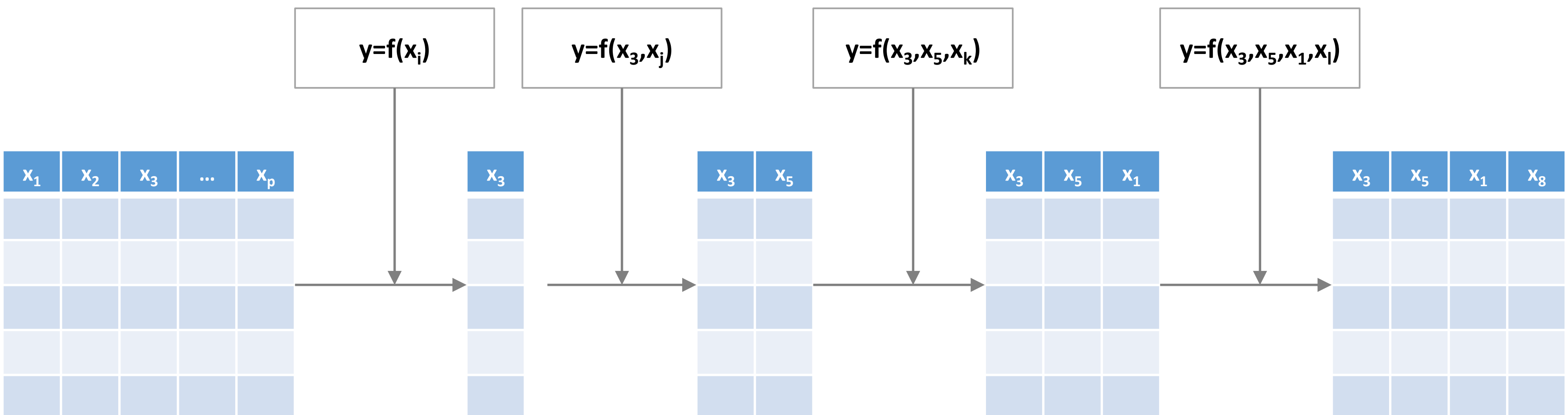
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_7 x_7 + \hat{\beta}_4 x_4 + \hat{\beta}_8 x_8, \quad R^2_{adj} = 0.75$$



- ✓ Final model: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_7 x_7 + \hat{\beta}_4 x_4, \quad R^2_{adj} = 0.76$

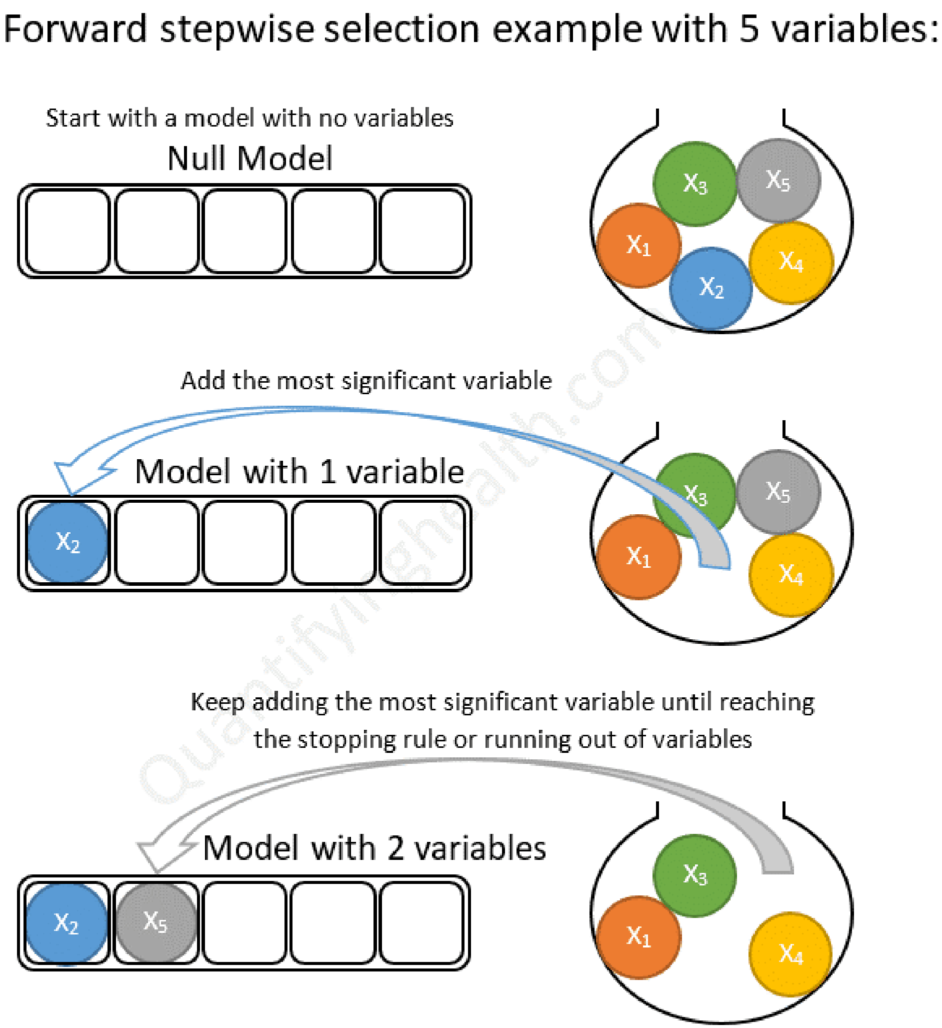
Forward Selection

- Forward selection
  - ✓ From the model with no variable, significant variables are sequentially added
  - ✓ Once the variable is selected, it will never be removed (The number of variables gradually increases)



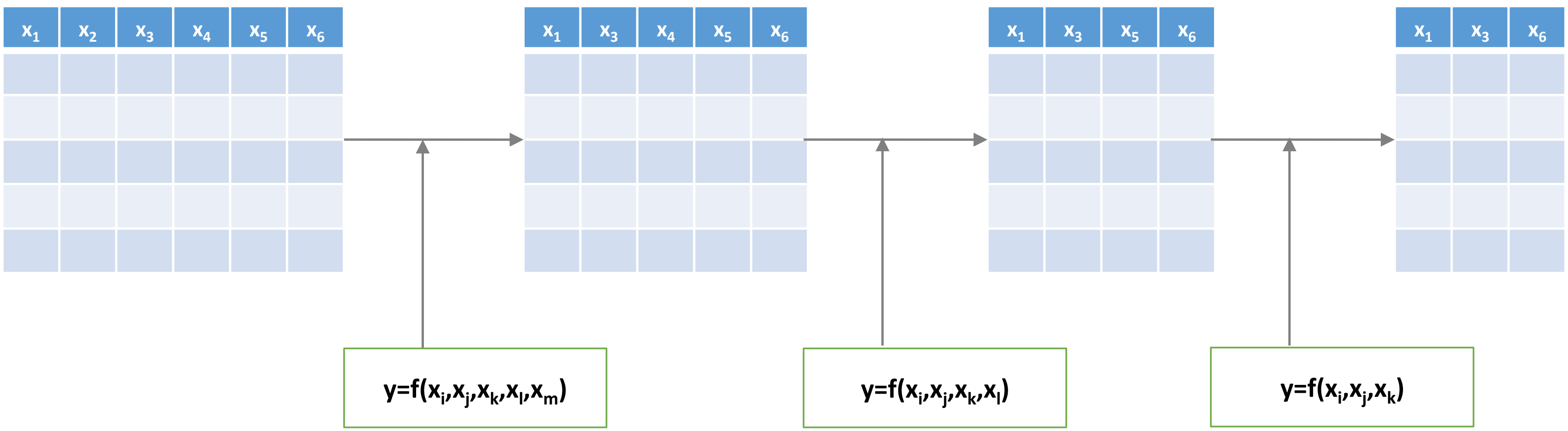
Forward Selection

- Illustrative Example



Backward Elimination

- Backward Elimination
  - ✓ From the model with all variables, irrelevant variables are sequentially removed
  - ✓ Once a variable is removed, it will never be selected (The number of variables gradually decreases)



Backward Elimination

- Backward Elimination Example
  - ✓ Backward elimination in the multiple linear regression
  - ✓ Begins with the model with all variable

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_6 x_6 + \hat{\beta}_7 x_7 + \hat{\beta}_8 x_8, \quad R^2_{adj} = 0.77$$

Backward Elimination

- Backward Elimination Example: Linear Regression
  - ✓ Remove the most irrelevant variable

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_6 x_6 + \hat{\beta}_7 x_7 + \hat{\beta}_8 x_8, \quad R^2_{adj} = 0.77$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_6 x_6 + \hat{\beta}_7 x_7 + \hat{\beta}_8 x_8, \quad R^2_{adj} = 0.65$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_6 x_6 + \hat{\beta}_7 x_7 + \hat{\beta}_8 x_8, \quad R^2_{adj} = 0.60$$

$$\boxed{\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_6 x_6 + \hat{\beta}_7 x_7 + \hat{\beta}_8 x_8, \quad R^2_{adj} = 0.77}$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_5 x_5 + \hat{\beta}_6 x_6 + \hat{\beta}_7 x_7 + \hat{\beta}_8 x_8, \quad R^2_{adj} = 0.62$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4 + \hat{\beta}_6 x_6 + \hat{\beta}_7 x_7 + \hat{\beta}_8 x_8, \quad R^2_{adj} = 0.73$$

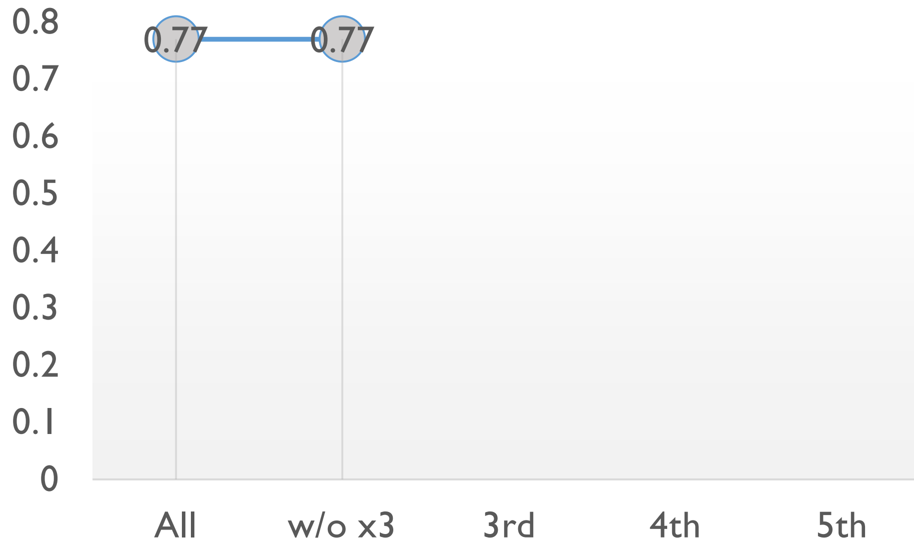
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_7 x_7 + \hat{\beta}_8 x_8, \quad R^2_{adj} = 0.71$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_6 x_6 + \hat{\beta}_8 x_8, \quad R^2_{adj} = 0.61$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_6 x_6 + \hat{\beta}_7 x_7, \quad R^2_{adj} = 0.74$$

Backward Elimination

- Backward Elimination Example: Linear Regression
  - ✓ Remove the most irrelevant variable



Backward Elimination

- Backward Elimination Example: Linear Regression
  - ✓ Remove the second most irrelevant variable

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \qquad\qquad + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_6 x_6 + \hat{\beta}_7 x_7 + \hat{\beta}_8 x_8, \quad R^2_{adj} = 0.77$$

$$\hat{y} = \hat{\beta}_0 \qquad\qquad + \hat{\beta}_2 x_2 + \qquad\qquad \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_6 x_6 + \hat{\beta}_7 x_7 + \hat{\beta}_8 x_8, \quad R^2_{adj} = 0.63$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 \qquad\qquad + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_6 x_6 + \hat{\beta}_7 x_7 + \hat{\beta}_8 x_8, \quad R^2_{adj} = 0.59$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \qquad\qquad + \hat{\beta}_5 x_5 + \hat{\beta}_6 x_6 + \hat{\beta}_7 x_7 + \hat{\beta}_8 x_8, \quad R^2_{adj} = 0.61$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \qquad\qquad + \hat{\beta}_4 x_4 \qquad\qquad + \hat{\beta}_6 x_6 + \hat{\beta}_7 x_7 + \hat{\beta}_8 x_8, \quad R^2_{adj} = 0.70$$

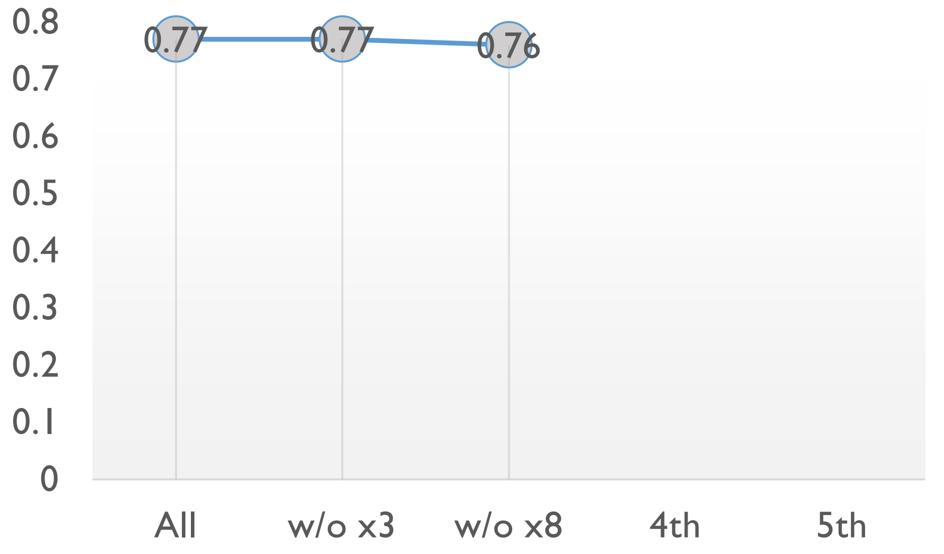
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \qquad\qquad + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 \qquad\qquad + \hat{\beta}_7 x_7 + \hat{\beta}_8 x_8, \quad R^2_{adj} = 0.69$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \qquad\qquad + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_6 x_6 \qquad\qquad + \hat{\beta}_8 x_8, \quad R^2_{adj} = 0.60$$

$$\boxed{\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \qquad\qquad + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_6 x_6 + \hat{\beta}_7 x_7 \qquad\qquad , \quad R^2_{adj} = 0.76}$$

Backward Elimination

- Backward Elimination Example: Linear Regression
  - ✓ Remove the most irrelevant variable



Backward Elimination

- Backward Elimination Example: Linear Regression
  - ✓ Remove the second most irrelevant variable

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_6 x_6 + \hat{\beta}_7 x_7 \quad , \quad R^2_{adj} = 0.76$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_6 x_6 + \hat{\beta}_7 x_7 \quad , \quad R^2_{adj} = 0.62$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_6 x_6 + \hat{\beta}_7 x_7 \quad , \quad R^2_{adj} = 0.58$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_5 x_5 + \hat{\beta}_6 x_6 + \hat{\beta}_7 x_7 \quad , \quad R^2_{adj} = 0.60$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_4 x_4 + \hat{\beta}_6 x_6 + \hat{\beta}_7 x_7 \quad , \quad R^2_{adj} = 0.66$$

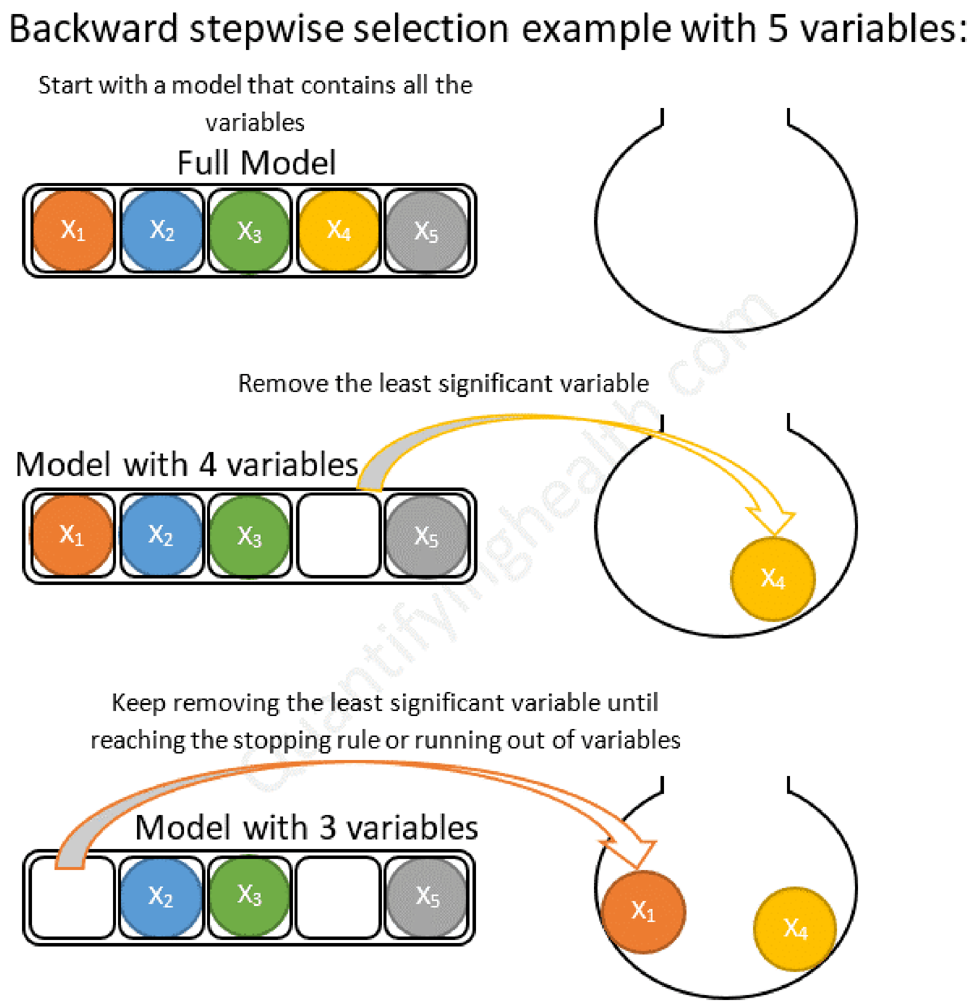
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_7 x_7 \quad , \quad R^2_{adj} = 0.67$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_6 x_6 \quad , \quad R^2_{adj} = 0.59$$

No variable is eliminated.

Backward Elimination

- Illustrative Example



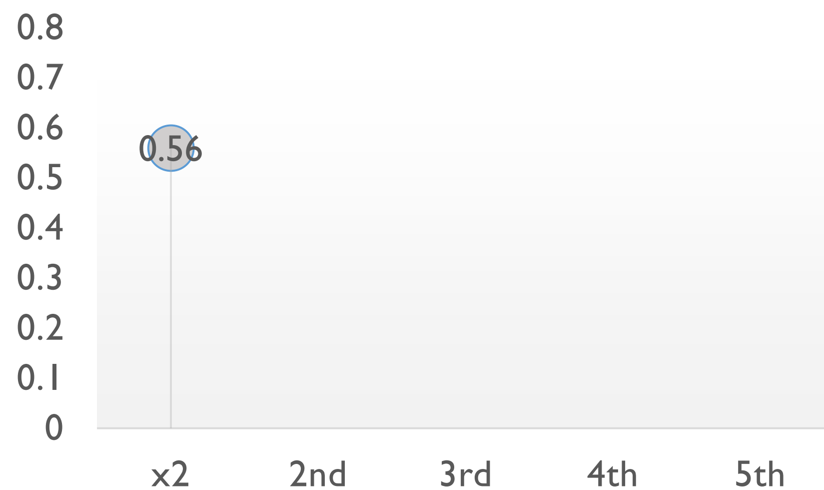
Stepwise Selection

- Stepwise Selection
  - ✓ From the model with no variable, conduct the forward selection and backward elimination alternately
  - ✓ Takes longer time than forward selection/backward elimination, but has more chances to find the optimal set of variables
  - ✓ Variables that is either selected/removed can be reconsidered for selection/removal
  - ✓ The number of variables increases in the early period, but it can either increase or decrease

Stepwise Selection

- Stepwise Selection Example: Linear Regression
  - ✓ Step 1: Select the most relevant variable

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2, \quad R^2_{adj} = 0.56$$



Stepwise Selection

- Stepwise Selection Example: Linear Regression
  - ✓ Step 2: Remove the most irrelevant variable

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2, \quad R^2_{adj} = 0.56$$

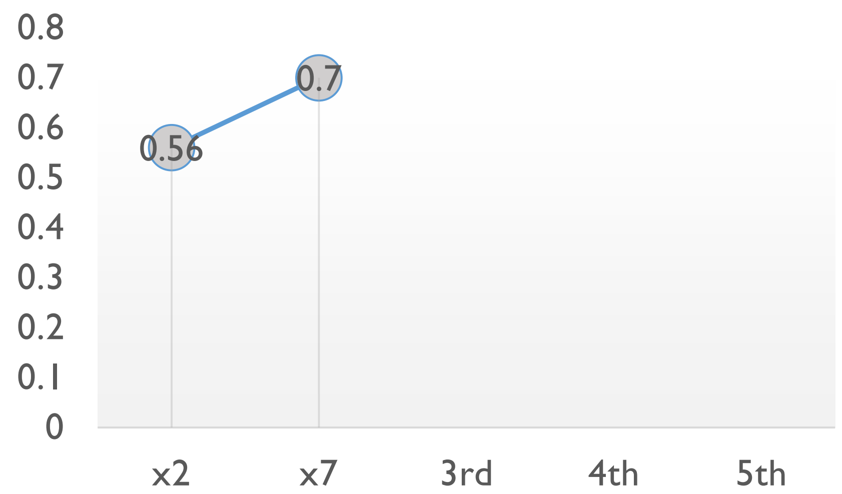
No variable is eliminated.

Stepwise Selection

- Stepwise Selection Example: Linear Regression
  - ✓ Back to the Step 1: Select the second most relevant variable

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2, \quad R^2_{adj} = 0.56$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_7 x_7, \quad R^2_{adj} = 0.70$$



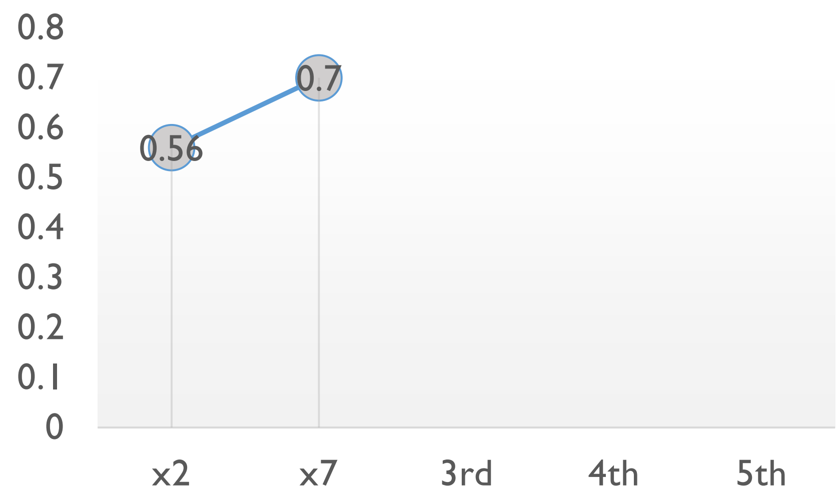
Stepwise Selection

- Stepwise Selection Example: Linear Regression
  - ✓ Back to the Step 2: Remove the most irrelevant variable

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2, \quad R^2_{adj} = 0.56$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_7 x_7, \quad R^2_{adj} = 0.70$$

No variable is eliminated.



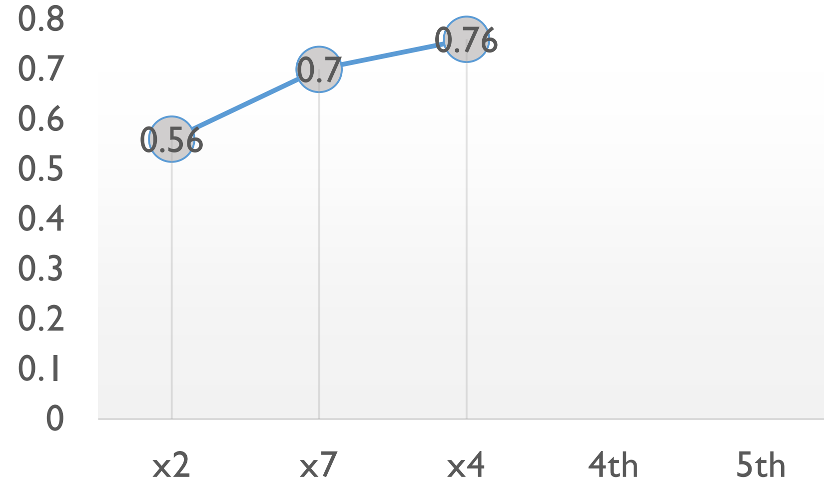
Stepwise Selection

- Stepwise Selection Example: Linear Regression
  - ✓ Back to the Step 1: Select the third most relevant variable

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2, \quad R^2_{adj} = 0.56$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_7 x_7, \quad R^2_{adj} = 0.70$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_7 x_7 + \hat{\beta}_4 x_4, \quad R^2_{adj} = 0.76$$



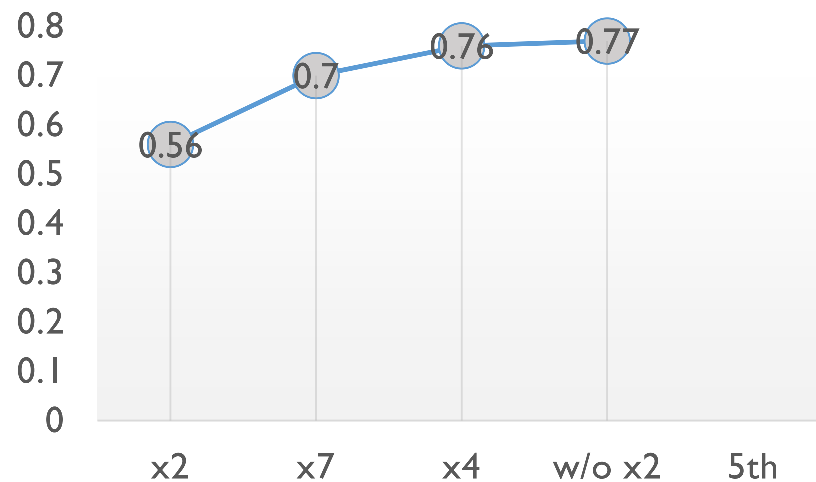
Stepwise Selection

- Stepwise Selection Example: Linear Regression
  - ✓ Back to the Step 2: Remove the most irrelevant variable

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2, \quad R^2_{adj} = 0.56$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_7 x_7, \quad R^2_{adj} = 0.70$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_7 x_7 + \hat{\beta}_4 x_4, \quad R^2_{adj} = 0.76$$



$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_4 x_4, \quad R^2_{adj} = 0.58$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_7 x_7, \quad R^2_{adj} = 0.70$$

$$\boxed{\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_4 + \hat{\beta}_7 x_7, \quad R^2_{adj} = 0.77}$$

The case which cannot be considered when x2 is selected in the forward selection method

Stepwise Selection

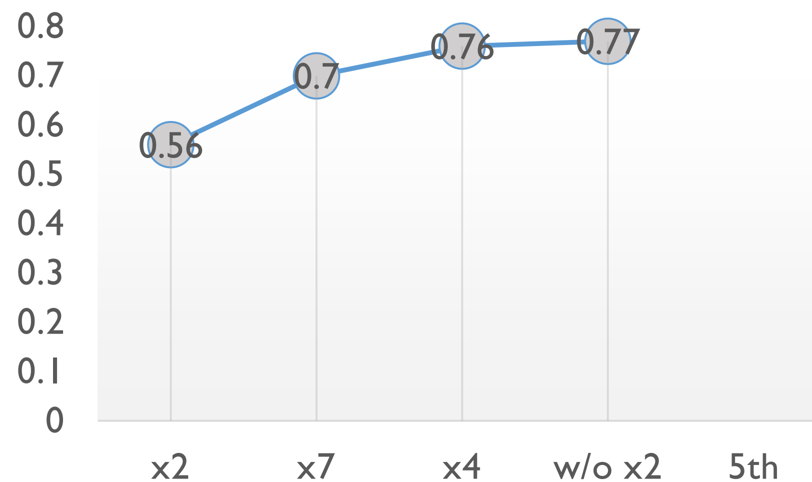
- Stepwise Selection Example: Linear Regression
  - ✓ Back to the Step 2: Remove the most irrelevant variable

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2, \quad R^2_{adj} = 0.56$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_7 x_7, \quad R^2_{adj} = 0.70$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_7 x_7 + \hat{\beta}_4 x_4, \quad R^2_{adj} = 0.76$$

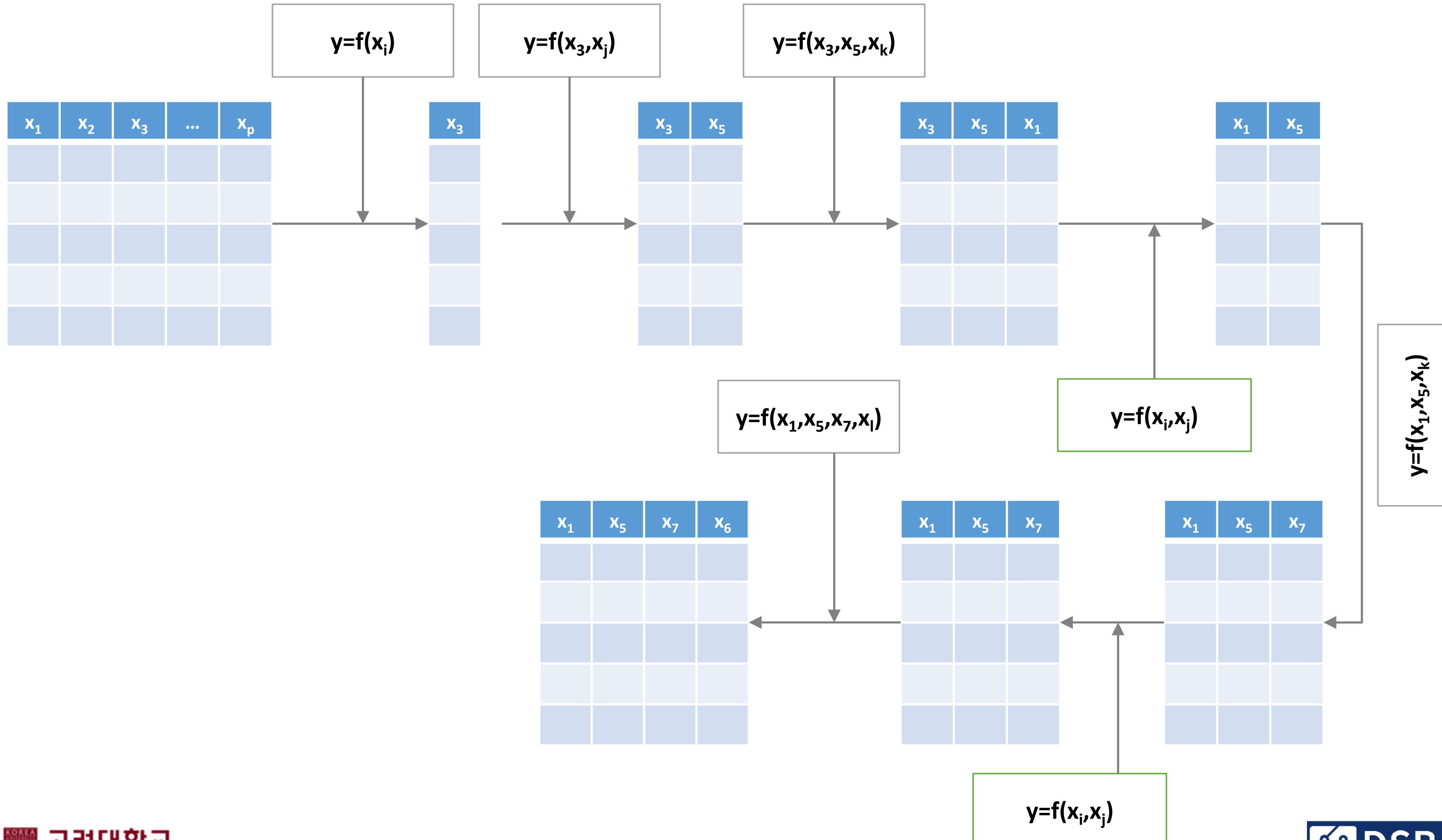
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_4 + \hat{\beta}_7 x_7, \quad R^2_{adj} = 0.77$$



✓ Repeat Step 1 & 2 until no variable is either selected or eliminated.

Stepwise Selection

- Stepwise selection example



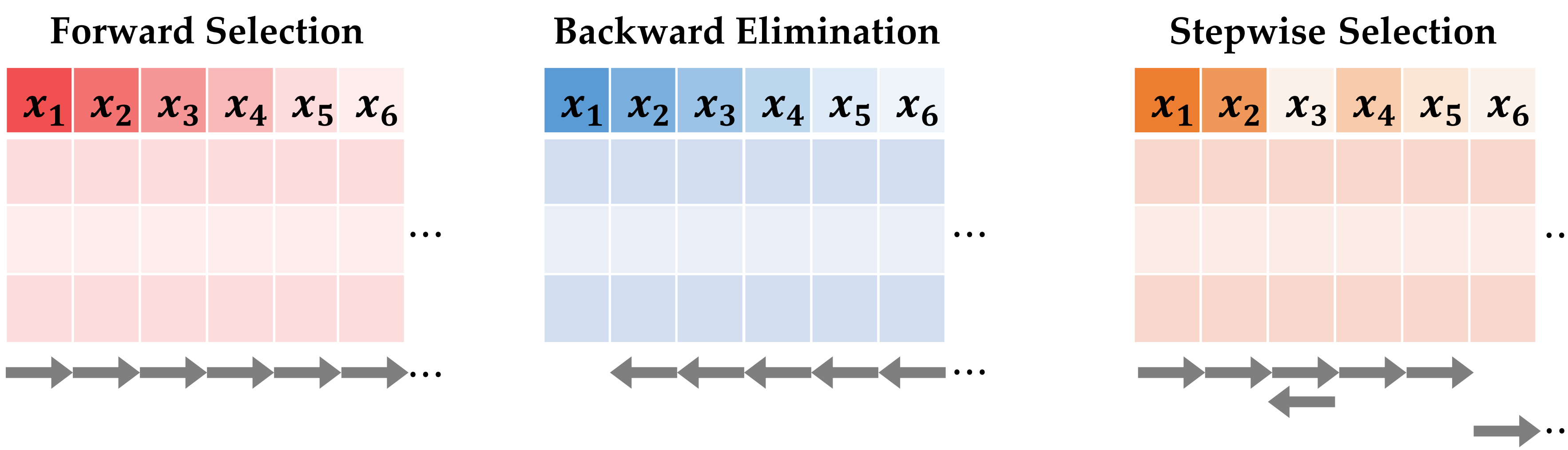
Stepwise Selection

- Stepwise Selection
  - ✓ Stepwise selection process

- ▶ Start with model with no predictors.
- ▶ Add variable with largest F-statistic (provided P less than some cut-off).
- ▶ Refit with this variable added. Recompute all F statistics for adding one of the remaining variables and add variable with largest F statistic.
- ▶ At each step after adding a variable try to eliminate any variable not significant at some level (that is, do BACKWARD elimination till that stops).
- ▶ After doing the backwards steps take another FORWARD step.
- ▶ Continue until every remaining variable is significant at cut-off level and every excluded variable is insignificant OR until variable to be added is same as last deleted variable.

Comparison among FS/BE/SS

- Illustrative Example



Performance Metrics

- Akaike Information Criteria (AIC)
  - ✓ Sum of squared error (SSE) with the number of variables as a penalty term

$$AIC = n \cdot ln\left(\frac{SSE}{n}\right) + 2k$$

- Bayesian Information Criteria (BIC)
  - ✓ SSE, number of variables, standard deviation obtained by the model with all variables

$$BIC = n \cdot ln\left(\frac{SSE}{n}\right) + \frac{2(k+2)n\sigma^2}{SSE} - \frac{2n^2\sigma^4}{SSE^2}$$

Performance Metrics

- Adjusted R^2
  - ✓ Simple R^2 increases when the number of variable increases

$$\begin{aligned} &\text{Model 1}: \quad y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \epsilon \\ &\text{Model 2}: \quad y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \ldots + \beta_{k+m} x_{k+m} \epsilon \end{aligned}$$

$$R^2(M2) \geq R^2(M1)$$

  - ✓ Use the adjusted R^2 that account for the number of variables (k)

$$\text{Adjusted } R^2 = 1 - \left(\frac{n-1}{n-k-1}\right)(1 - R^2) = 1 - \frac{n-1}{n-k-1}\frac{SSE}{SST}$$

고려대학교 KOREA UNIVERSITY

DSBA Data Science & Business Analytics


Q & A

References

Other materials

- Figures in the front page: https://medium.com/@arora.nishank91/feature-selection-for-faster-analytics-70a56132349e