1. 왜 변수가 많아지면 R^2이 무조건 상승하는가?

$\min_\beta \sum_{i=1}^{n} \varepsilon_i^2 = \min_\beta \sum_{i=1}^{n} (y_i - X_i\beta)^2$.

The above equation solves for the values of the coefficients such that the squared errors are minimized, or equivalently, for the values of the coefficients such that what you are able to explain, i.e. the R^2, is **maximized**.

Therefore, whenever you add a variable to your model, the value of its estimated coefficient can either be zero, in which case the proportion of explained variance (R^2) stays unchanged, or take a nonzero value **because it improves the quality of the fit**. By construction, your R^2 *cannot* be smaller after adding a variable.

2. R^2(결정계수)보다 RSE가 더 쓸모있는 상황이 있나요? 0

RSE: 130만 루피 R^2: 0.89

Qualitative Predictors

> 어떤 변수들은 범주형이다. (qualitative: 정성적)

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i & \text{if } i\text{th person is female} \\ \beta_0 + \epsilon_i & \text{if } i\text{th person is male.} \end{cases} \quad (3.27)$$

위와 같이 회귀식에 포함시킨다.

그 성질이 3가지인 경우는 이런 식

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i & \text{if } i\text{th person is Asian} \\ \beta_0 + \beta_2 + \epsilon_i & \text{if } i\text{th person is Caucasian} \\ \beta_0 + \epsilon_i & \text{if } i\text{th person is African American.} \end{cases} \quad (3.30)$$

이런 것들을 dummy variable이라고 한다. 이는 범주형 변수를 다루는 한가지 방법이고

Convert to Number Combine Levels 등이 있다.

메인 포인트는 범주를 정량화시키는 것!

Interaction effect

변수 x1과 x2가 가지는 시너지 효과가 있을 수 있다.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon. \qquad (3.31)$$

그래서 다음과 같이 interaction term를 넣어주면 그러한 시너지 효과를 모델에 반영 할 수 있다.

하지만!

the associated main effects (in this case, TV and radio) do not. The *hierarchical principle* states that *if we include an interaction in a model, we should also include the main effects, even if the p-values associated with their coefficients are not significant.* In other words, if the interaction be-

그러므로 interaction term을 넣으면 그와 대응되는 변수들도 함께 넣어주어야 된다.

왜..?

Why do we have the hierarchical principle in adding interactions to a model? What is the significance of it?

Peter Flom

Independent statistical consultant for researchers in social, medical and behavioral sciences

Statistical analyst for researchers in the social, medical and behavioral sciences.

Learning disabled adult. Be sure to follow me on: Random thoughts (my Quora blog) ... (more)

Suppose, you are modeling the relationship between income (logged) as the DV and race/ethnic group (White/other) and sex (male/female) and their interaction (I am keeping this simple - I know there are more than 2 ethnicities and more than 2 sexes). You have four combinations:

White men
White women
Other men
Other women

If you include the interaction and the main effects, you are saying that each of these could be different from the other three. If you do not include the interaction, but only the main effects, you are saying that all four could be different, but that the difference between White men and White women is the same as between other men and other women; this might be reasonable (but I've seen evidence that it is not true)

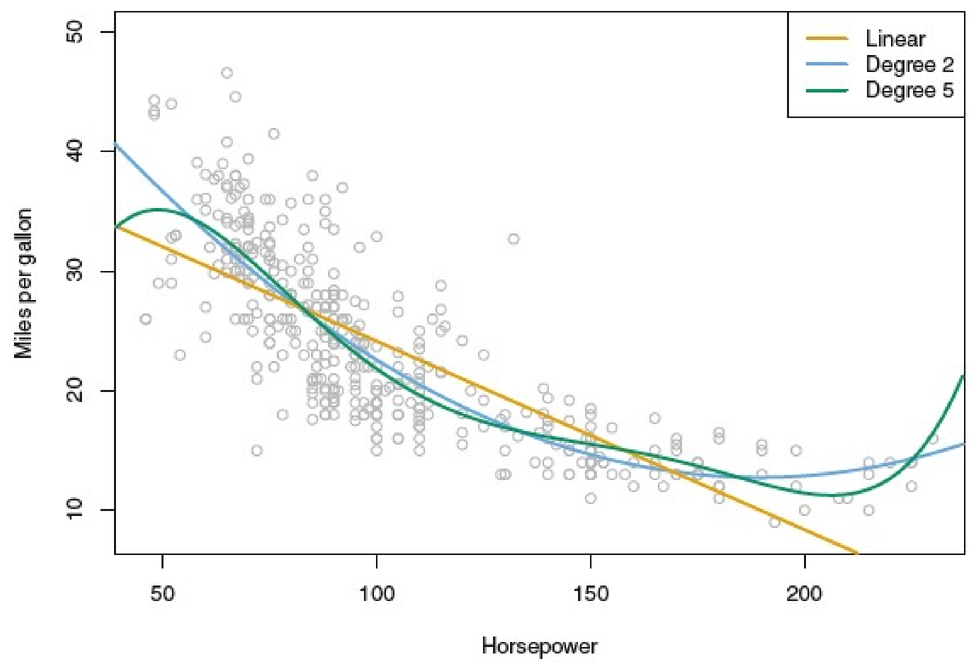
But if you only include the interaction, then you are saying that the first three are all equal and the last is different. That makes no sense here and it is hard to think of cases where it does make sense (but not quite impossible, see PsycNET - Display Record)

Hierarchical principle... 뭔가 중요해 보인다

Non-linear Relationship

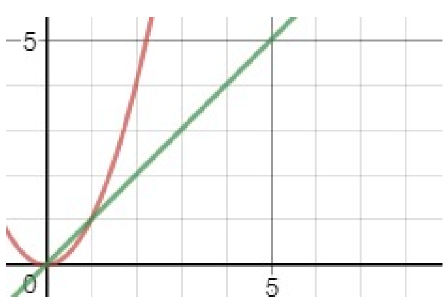
Polynomial regression을 통해 비선형적인 관계도 모델에 반영할 수 있다.

3.3 Other Considerations in the Regression Model 91



$$\texttt{mpg} = \beta_0 + \beta_1 \times \texttt{horsepower} + \beta_2 \times \texttt{horsepower}^2 + \epsilon \qquad (3.36)$$

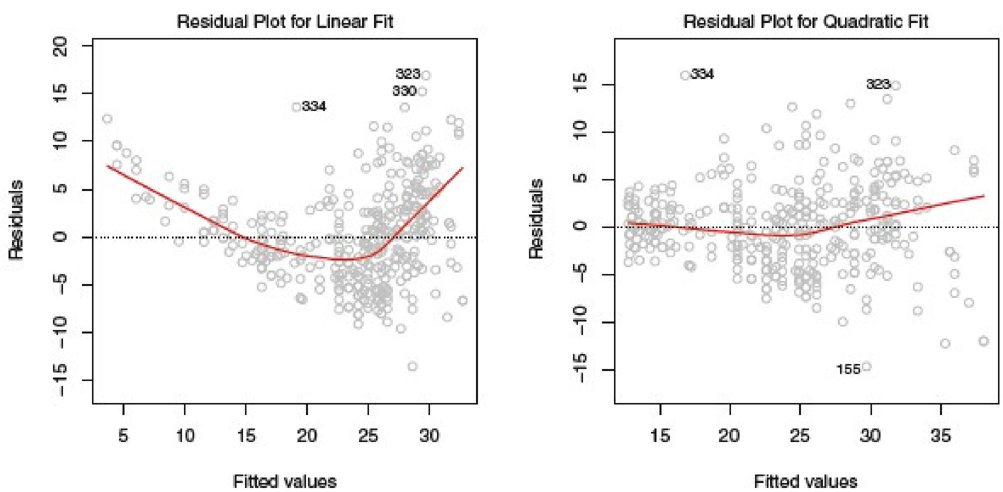
이런식으로 비선형적 관계를 모델에 반영할 수 있다.
하지만 결국 이것도 선형적인 모델, horsepower대신에 horsepower^2라는 새로운 변수를 이용하는 것



$y = x^2$ -> $y = x$

Potential Problems

1. Non-linearity of the response-predictor relationships



그렇다면 어떤 식으로 polynomial 화 시켜야 될까? $\log X$, $\sqrt{X}$, and X^2,

2. Correlation of error terms

An important assumption of the linear regression model is that the error terms, $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$, are uncorrelated. What does this mean? For instance,

are based on the assumption of uncorrelated error terms. If in fact there is correlation among the error terms, then the estimated standard errors will tend to underestimate the true standard errors. As a result, confi-

이는

1. error terms 안에 우리가 반영하지 못한 변수들이 숨어있다는 것.

2. Functionally misspecified model(비선형 데이터를 선형 모델로 설명하려 할 때)

3. Measurement error in independent variable (1m 자로 벼룩 키재기) 등의 이유로 생겨날 수 있다.

Autocorrelation

https://www.youtube.com/watch?v=jt5nl2VEpwg

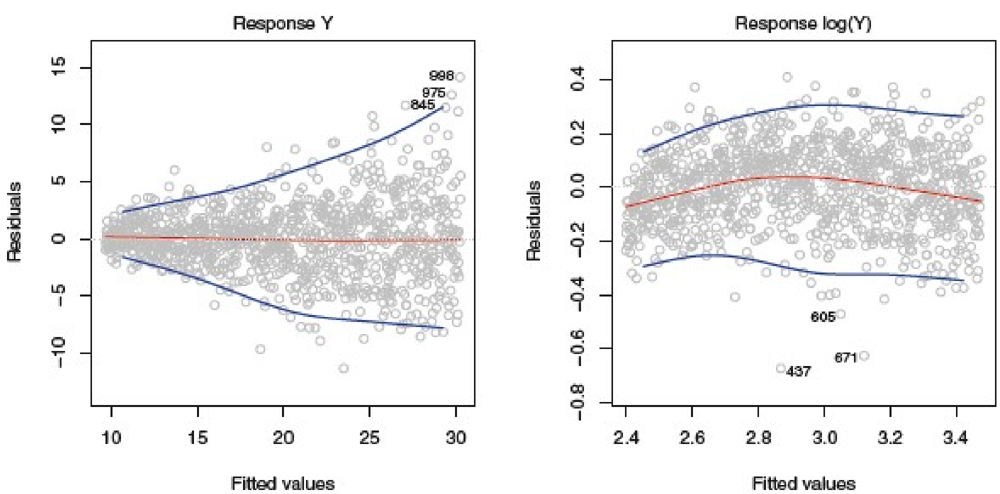
- $y_i = f(x_i) + \varepsilon_i$
- $\varepsilon_i = \rho\varepsilon_{i-1} + u_i,\ u_i \sim N(0, \sigma^2)$ iid
- ρ = autocorrelation parameter, $|\rho| < 1$
  - $\rho = 0$ means no autocorrelation
  - $\rho > 0$ means positive autocorrelation ("persistence")

- Properties of the error terms:
  - $E[\varepsilon_i] = 0$
  - $var[\varepsilon_i] = \sigma^2/(1-\rho^2)$
  - $cov[\varepsilon_i, \varepsilon_{i-1}] = \rho\sigma^2/(1-\rho^2) = \rho var[\varepsilon_i]$
  - $cov[\varepsilon_i, \varepsilon_{i-s}] = \rho^s\sigma^2/(1-\rho^2) = \rho^s var[\varepsilon_i]$, for $s > 0$

이게 정확히 어떤 영향을 미치는 걸까?

3. Non-constant variance of error terms

$\mathrm{Var}(\epsilon_i) = \sigma^2$. *heteroscedasticity,*



the response Y using a concave function such as $\log Y$ or $\sqrt{Y}$.

4. **Outliers**

outlier causes the R^2 to decline from 0.892 to 0.805.

this problem, instead of plotting the residuals, we can plot the *studentized residuals*, computed by dividing each residual e_i by its estimated standard error. Observations whose studentized residuals are greater than 3 in absolute value are possible outliers. In the right-hand panel of Figure 3.12, the

5. **High-leverage points**

unusual value for x_i.

4와 5의 콤비네이션은 특히 위험!(스티브 잡스 같은 데이터 포인트들)

$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i'=1}^{n}(x_{i'} - \bar{x})^2}.$$

6. **Collinearity**

Collinearity refers to the situation in which two or more predictor variables are closely related to one another. The concept of collinearity is illustrated

Since collinearity reduces the accuracy of the estimates of the regression coefficients, it causes the standard error for $\hat{\beta}_j$ to grow. Recall that the

A simple way to detect collinearity is to look at the correlation matrix of the predictors. An element of this matrix that is large in absolute value

multicollinearity. -> *variance inflation factor* (VIF).

$$\mathrm{VIF}(\hat{\beta}_j) = \frac{1}{1 - R^2_{X_j|X_{-j}}},$$

where $R^2_{X_j|X_{-j}}$ is the R^2 from a regression of X_j onto all of the other predictors. If $R^2_{X_j|X_{-j}}$ is close to one, then collinearity is present, and so the VIF will be large.

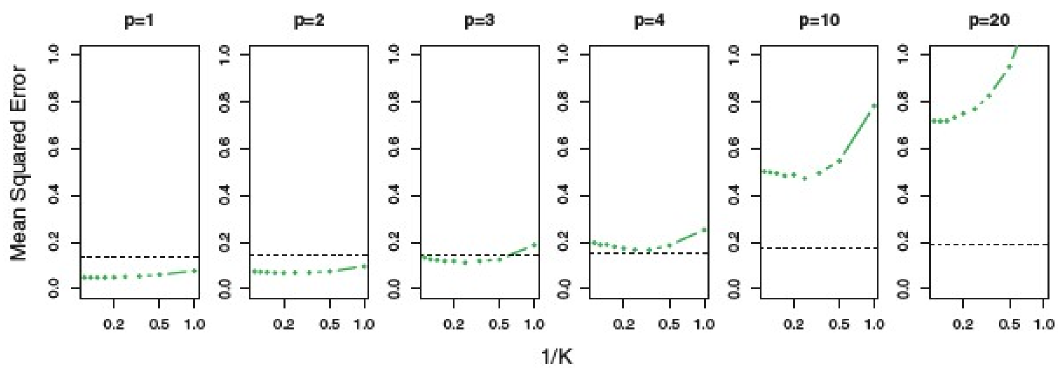
해당 variable를 삭제하거나 아니면 연관 있는 variable들을 모아 하나의 새로운 variable를 만들던가!

Comparison of Linear Regression with K-Nearest Neighbors

(KNN은 노답상황에서 사용하는 기법)

The answer is simple: *the parametric approach will outperform the non-parametric approach if the parametric form that has been selected is close to the true form of f.* Figure 3.17 provides an example with data generated

KNN. This decrease in performance as the dimension increases is a common problem for KNN, and results from the fact that in higher dimensions there is effectively a reduction in sample size. In this data set there are 100 training observations; when $p = 1$, this provides enough information to accurately estimate $f(X)$. However, spreading 100 observations over $p = 20$ dimensions results in a phenomenon in which a given observation has no *nearby neighbors*—this is the so-called *curse of dimensionality*. That is,



다음주까지 한번 더 조사해보면 좋을 것들

Hierarchical principle

Variance and bias tradeoff

Curse of dimensionality

Potential problems의 수학적 증명