정 규 세 션 2 주 차

ToBig's 11기 심은선

Regression Analysis

회귀분석

Contents

머신러닝이란?

"만약 컴퓨터 프로그램이 특정한 태스크 T를 수행할 때 성능 P 만큼 개선되는 경험 E를 보이면 그 컴퓨터 프로그램은 태스크와 성능 P에 대해 경험 E를 학습했다라고 할 수 있다."

-Tom Mitchell

Ex) 선형회귀, KNN, SVM, Decision Tree, PCA 등

머신러닝 알고리즘 종류

1) 지도학습

- 데이터의 레이블(정답)이 주어진 상태에서 컴퓨터를 학습

  ex) 회귀분석, 로지스틱회귀, KNN

2) 비지도학습

- 데이터의 레이블이 주어지지 않은 상태에서 컴퓨터를 학습

  >>데이터의 숨겨진 특성 파악

  ex) Clustering

3) 강화학습

- 에이전트가 주어진 환경(state)에서 행동(action)을 취하고 이로부터 보상(reward)을 얻는데, 보상을 최대화하는 방향으로 학습

레이블?

-저번주 EDA과제의 Hammer_price변수!

(=종속변수)

road_name	road_bunji1	road_bunji2	Close_date	Close_result	point.y	point.x	Hammer_price
해운대해변로	30.0	NaN	2018-06-14 00:00:00	배당	35.162717	129.137048	760000000
마린시티2로	33.0	NaN	2017-03-30 00:00:00	배당	35.156633	129.145068	971889999
모라로110번길	88.0	NaN	2017-12-13 00:00:00	배당	35.184601	128.996765	93399999
황령대로319번가길	110.0	NaN	2017-12-27 00:00:00	배당	35.154180	129.089081	256899000
오작로	51.0	NaN	2016-10-04 00:00:00	배당	35.099630	128.998874	158660000

머신러닝 알고리즘 종류

머신러닝 잘하는 방법???

No free lunch

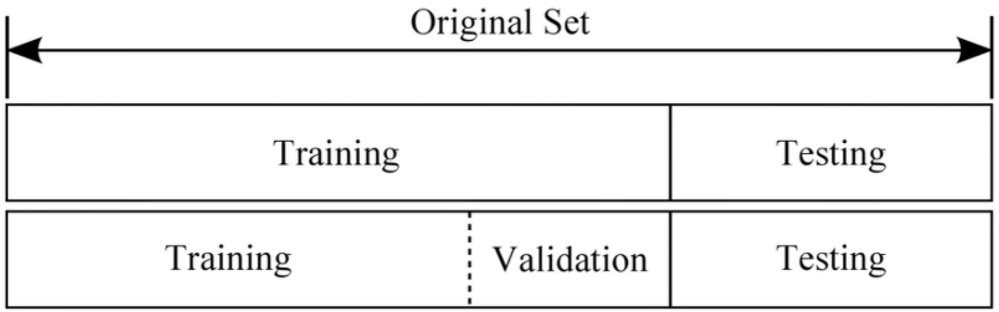
EDA, 전처리부터 적절한 모델, 파라미터 등등… 많이 고민해보고 직접 해봐야함!

<유용한 자료>

https://github.com/ExcelsiorCJH/Hands-On-ML/blob/master/Chap02-End_to_End_ML_Project/Chap02-End_to_End_ML_Project.ipynb

https://github.com/KaggleBreak/walkingkaggle

Training and Testing



1) Training data

- 모델에 넣어 파라미터를 inference할 때 사용
- 현재, 과거의 경험(지식)

2) Test data

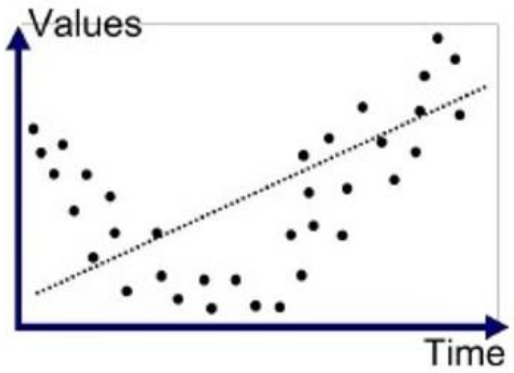
- inference한 모델이 실제 다른 데이터에도 잘 맞는지 평가
- 미래의 들어올 데이터를 imitate
- test data는 training 데이터와 관련 없어야 함

>> 데이터를 보통 Train:Test=7:3으로 분할해서 모델을 돌림

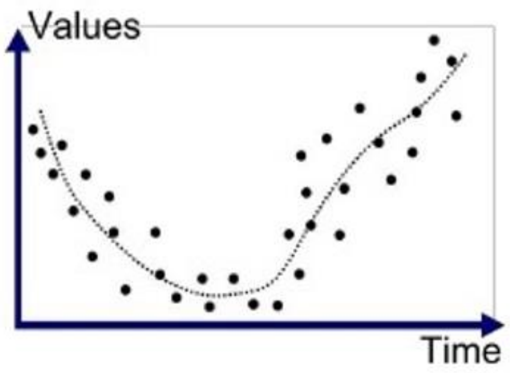
cf) Dev set을 사용할 경우 보통 Train:Dev:Test=6:2:2으로 분할 (Dev(validation) set은 하이퍼파라미터 튜닝 목적)

과대적합(Overfitting)과 과소적합(Underfitting)

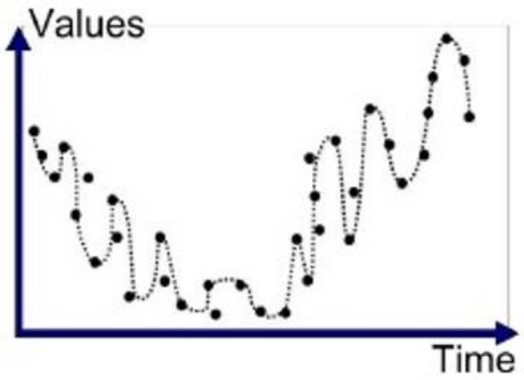
- 과대적합: train 데이터를 매우 잘 설명하나, 새로운 데이터가 들어오면 잘 맞추지 못함 (일반성이 떨어짐)
- 과소적합: 모델이 단순해서 train 데이터를 완벽하게 맞추지 못하지만, 새로운 데이터도 어느정도 잘 맞춤 (일반화된 모델)



Underfitted



Good Fit



Overfitted

Unit 02 | Bias & Variance

Bias and Variance

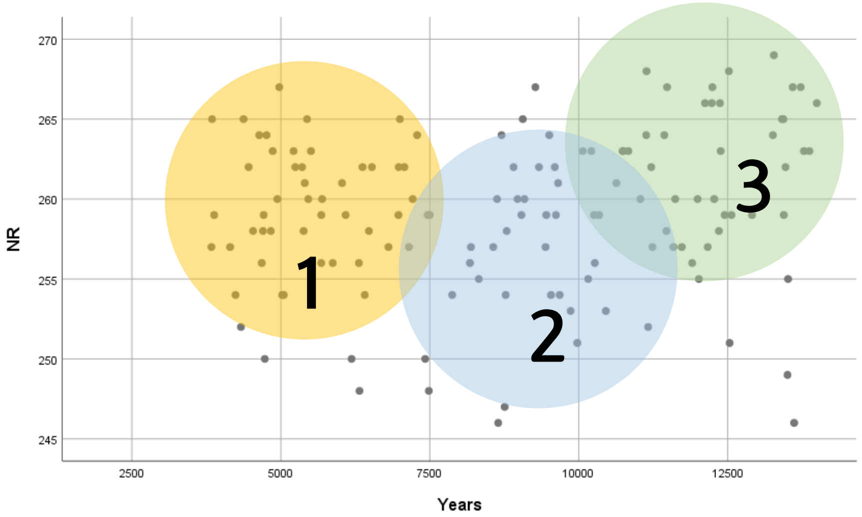
1) Bias

- True function과 모든 데이터셋으로 얻는 머신러닝 function과의 차이
  (머신러닝 모델을 사용하기 때문에 필연적으로 발생하는 오차)

2) Variance

- 모든 데이터셋으로 얻는 ML function과
  수집된 일부 데이터로 얻은 ML function의 차이
  (제한된 데이터로 인해 발생하는 오차)

>>Bias, Variance는 오차이니까 작으면 좋음!



Bias and Variance

<모델의 정확도 높임>

: 모델이 True function과 가까워져 Bias **감소**, 주어진 데이터에 과하게 맞추기 때문에 Variance **증가**

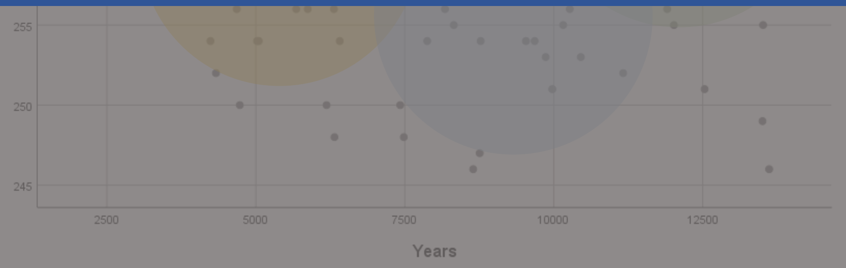
<모델의 일반화>

: True function과 멀어져 Bias **증가**, 다른 데이터가 주어져도 비슷하게 맞추기 때문에 Variance **감소**

(과대적합- Bias 작음, Variance 큼 vs. 과소적합- Bias 큼, Variance 작음)

\>> **Bias와 Variance의 Trade-off 관계**

(제한된 데이터로 인해 발생하는 오차)

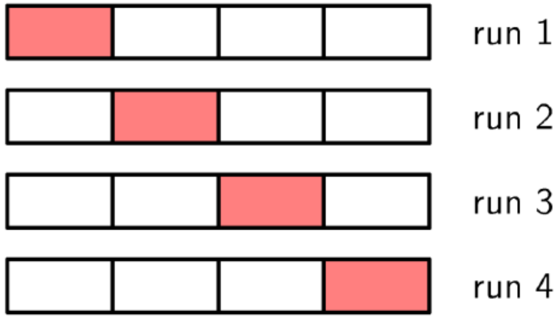


K-fold Cross validation(교차검증)

- 데이터가 적은 경우에 데이터를 분할하면, 검증용 데이터가 적어서 검증의 신뢰도가 떨어진다.
  그렇다고 검증용 데이터를 늘리면 train데이터가 적어져 학습이 제대로 되지 않는다.

>> 한정된 데이터에서 데이터가 많은 것처럼 흉내냄

ex)4-fold cross validation



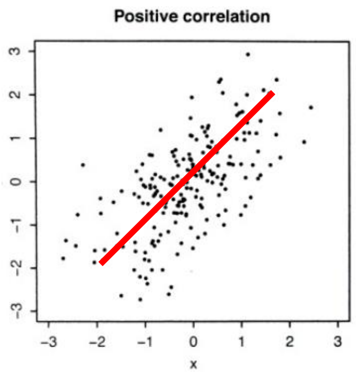
전체 데이터를 4개로 분할해서
3개를 train에 사용하고, 1개로 검증(test).
4개의 모델의 평균을 최종 성능 또는
최종 모델의 하이퍼파라미터로 사용

상관계수

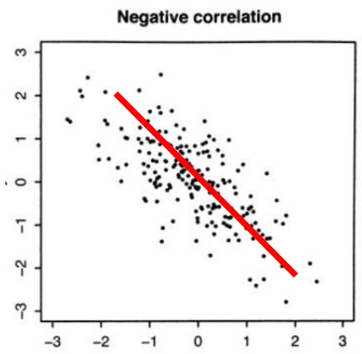
$$r=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\cdot\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

공분산을 두 변수의 표준편차로 나눔

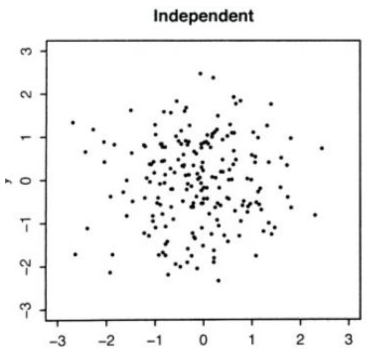
- 두 변수간의 선형관계를 분석(인과관계x)
- (-1) ~ 1의 값을 가진다
- 1에 가까우면 강한 양의 상관관계, 0이면 무상관, -1에 가까우면 강한 음의 상관관계
- 선형관계만 측정할 뿐 다른 관계(2차…)는 알 수 없음



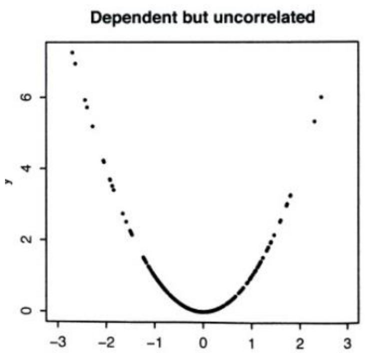
양의 상관



음의 상관



무상관



무상관

X가 증가할때 y가 증가(감소)하는 선형관계가 아니므로 (2차이지만) 무상관

회귀분석

- 설명변수의 선형결합(1차식)으로 종속변수를 설명하는 분석방법
  ex)기온(x)에 따른 빙수판매량(y)
- 설명변수의 개수에 따라 설명변수가 한 개면 단순회귀분석, 두 개 이상이면 다중회귀분석
- 데이터를 회귀모델에 돌리는 것(training)은 회귀계수($\beta_0, \beta_1, \beta_2 \cdots$)를 찾는 것

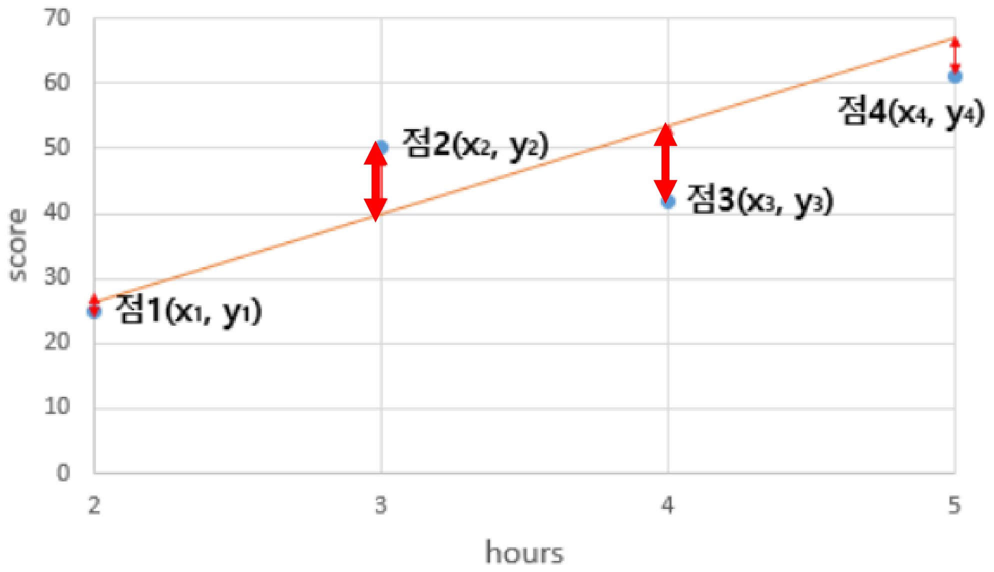
$$y = \beta_0 + \beta_1 x + \varepsilon_i$$

<단순선형회귀>

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots. + \beta_n X_n + \varepsilon$$

<다중선형회귀>

최소제곱법(Least squares)



회귀식이 예측한 값과 실제 값이 가까울 수록 좋으니까
그 차이를 최소화하자!

최소제곱법
: 실제 종속변수와 회귀식이 예측한 종속변수의 차이(오차)의 제곱합을 최소화하는 알고리즘

$$L = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

Loss function

$$\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip})^2$$

최소제곱법(Least squares)

- 목적함수(loss function) (단순회귀)

Loss function $L = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$

최소제곱법
: 실제 종속변수와 회귀식이 예측한 종속변수의 차이(오차)의 제곱합을 최소화하는 알고리즘

최소화하므로 편미분=0

$$\frac{\partial L}{\partial \beta_0} = -2\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\frac{\partial L}{\partial \beta_1} = -2\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)x_i = 0$$

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}, \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

(β_1의 최소제곱 추정량)

최소제곱법(Least squares)

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

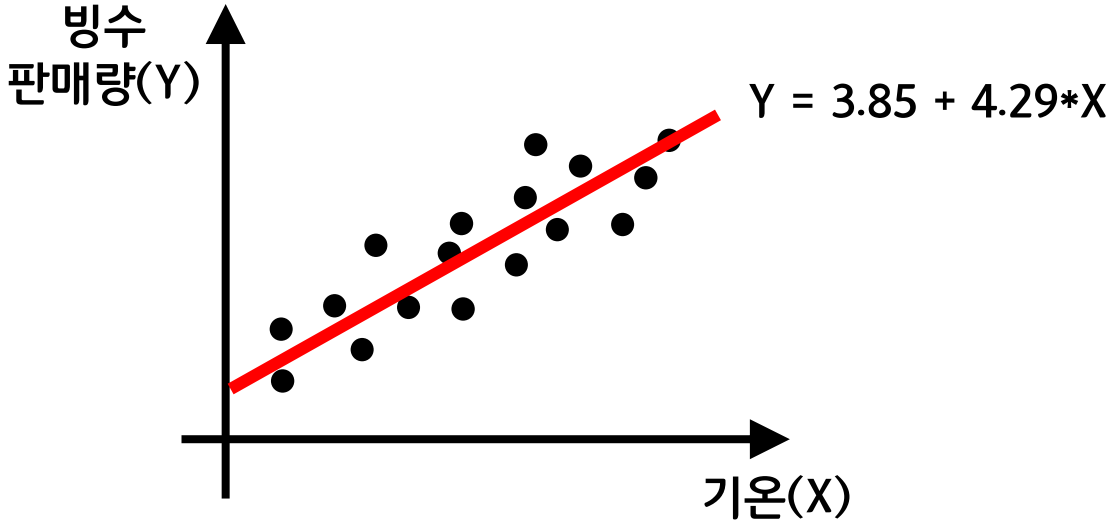
<적합된 회귀식>

-예측: x_i에 값을 대입해 종속변수 예측

-해석: $\hat{\beta}_0$ = intercept(절편)

$\hat{\beta}_1$ = x_i가 한 단위 증가할 때 종속변수의 증가(감소)량

-베타hat: 추정된 회귀계수



빙수판매량 = 3.85 + 4.29*(기온)

기온과 관계없이(기온이 0이어도) 빙수 3.85그릇이 팔리고
기온이 1도 증가하면 빙수 4.29그릇이 더 팔린다

최소제곱법(Least squares)-행렬

β_0 와 곱해져서 intercept를 만드는 열벡터

$$\mathbf{Y}=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},\quad \mathbf{X}=\begin{bmatrix} \boxed{\begin{matrix}1\\1\\\vdots\\1\end{matrix}} & \boxed{\begin{matrix}x_{11}\\x_{21}\\\vdots\\x_{n1}\end{matrix}} & \begin{matrix}\cdots\\\cdots\\\cdots\\\cdots\end{matrix} & \boxed{\begin{matrix}x_{1p}\\x_{2p}\\\vdots\\x_{np}\end{matrix}} \end{bmatrix},\quad \boldsymbol{\beta}=\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix},\quad \boldsymbol{\epsilon}=\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

피쳐1 … 피쳐n

(피쳐(feature): 변수)

road_name	road_bunji1	road_bunji2	Close_date	Close_result	point.y	point.x
해운대해변로	30.0	NaN	2018-06-14 00:00:00	배당	35.162717	129.137048
마린시티2로	33.0	NaN	2017-03-30 00:00:00	배당	35.156633	129.145068
모라로110번길	88.0	NaN	2017-12-13 00:00:00	배당	35.184601	128.996765
황령대로319번가길	110.0	NaN	2017-12-27 00:00:00	배당	35.154180	129.089081
오작로	51.0	NaN	2016-10-04 00:00:00	배당	35.099630	128.998874

최소제곱법(Least squares)-행렬

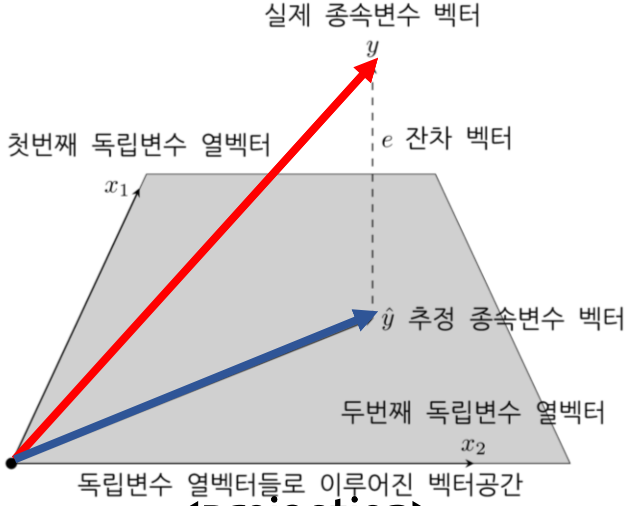
$$\text{Loss} = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip})^2$$
$$= (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta).$$

최소화하므로 편미분=0

$\Rightarrow \beta$에 관해 미분 $\quad \dfrac{\partial S}{\partial \beta} = -2X'Y + 2X'X\beta = 0$

$\rightarrow X'X\hat{\beta} = X'Y$

$\rightarrow \hat{\beta} = (X'X)^{-1}X'Y$



<projection>

추정된 회귀계수를 행렬을 통해
다음과 같이 표현할 수도 있음

Transpose(전치행렬)

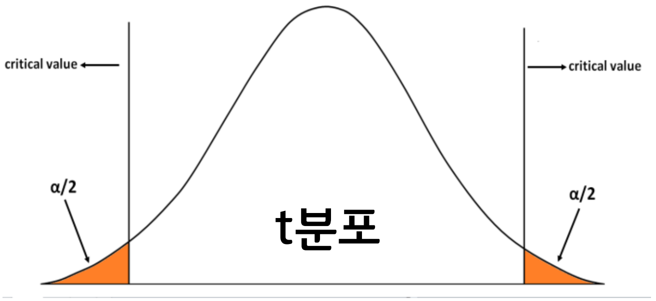
$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

적합된 회귀식: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip},$

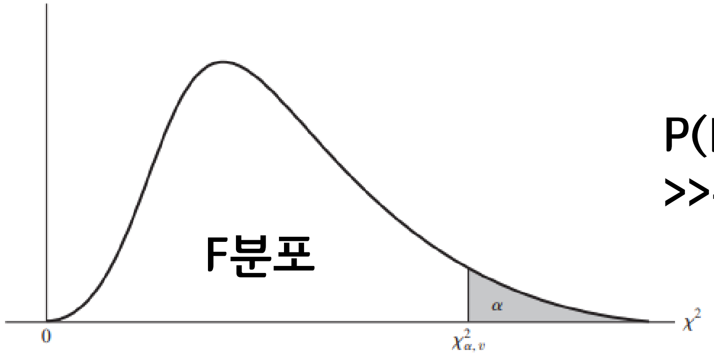
회귀분석 결과

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     median_house_value   R-squared:                       0.677
Model:                            OLS   Adj. R-squared:                  0.676
Method:                 Least Squares   F-statistic:                     2669.
Date:                Tue, 05 Jun 2018   Prob (F-statistic):               0.00
Time:                        14:52:54   Log-Likelihood:            -2.5564e+05
No. Observations:               20433   AIC:                         5.113e+05
Df Residuals:                   20416   BIC:                         5.114e+05
Df Model:                          16
Covariance Type:            nonrobust
==================================================================================================
                                 coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------------
Intercept                  -2.483e+06   8.79e+04    -28.267      0.000   -2.66e+06   -2.31e+06
ocean_proximity[T.INLAND]   -3.19e+04   1687.358    -18.906      0.000   -3.52e+04   -2.86e+04
ocean_proximity[T.ISLAND]   1.269e+05   2.94e+04      4.314      0.000    6.92e+04    1.85e+05
ocean_proximity[T.NEAR BAY] -7677.1895  1833.345     -4.188      0.000   -1.13e+04   -4083.687
ocean_proximity[T.NEAR OCEAN] -721.1056 1512.901     -0.477      0.634   -3686.513    2244.302
longitude                  -2.915e+04    992.666    -29.365      0.000   -3.11e+04   -2.72e+04
latitude                   -2.847e+04    984.297    -28.924      0.000   -3.04e+04   -2.65e+04
housing_median_age          1140.3596     42.411     26.888      0.000    1057.231    1223.488
total_rooms                 7416.5613   7599.836      0.976      0.329   -7479.726    2.23e+04
total_bedrooms                -1.3510      7.521     -0.180      0.857     -16.092      13.390
population                 -5958.6495   7421.124     -0.803      0.422   -2.05e+04    8587.348
households                    12.2682      8.324      1.474      0.141      -4.048      28.585
median_income               3.404e+04    582.505     58.445      0.000    3.29e+04    3.52e+04
income_cat                   1.42e+04   1063.823     13.344      0.000    1.21e+04    1.63e+04
rooms_per_household         4.526e+04   9363.233      4.834      0.000    2.69e+04    6.36e+04
bedrooms_per_room           3.922e+05   1.63e+04     24.114      0.000     3.6e+05    4.24e+05
population_per_household    -1.21e+05   1.05e+04    -11.537      0.000   -1.42e+05      -1e+05
==================================================================================================
Omnibus:                     4084.560   Durbin-Watson:                   1.050
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            13229.172
Skew:                           1.014   Prob(JB):                         0.00
Kurtosis:                       6.380   Cond. No.                     1.78e+05
==============================================================================
```

coef: 추정한 회귀계수 값
P>|t|: 추정된 회귀계수의 p값(유의확률, 양측)
P(F통계량): 회귀 모형 전체의 유의성 판단(단측)



p값 > 유의수준(보통 0.05)
>>통계적으로 의미없는 추정값
>>해당 변수 삭제/다른 조치
(p값이 클수록 안좋음)



P(F통계량) > 유의수준
>>통계적으로 의미없는 모형

변수선택법

1) 후진제거법(backward elimination)

: 독립변수 모두 사용한 모델에서 제거해도 큰 변화가 없는 독립변수를 하나씩 제거

2) 전진선택법(forward selection)

: 독립변수를 하나도 넣지 않은 모델에서 중요한 독립변수를 하나씩 추가

>>두 방법은 한번 빠진(선택한) 변수를 다시 추가(제거)할 수 없음

3) 단계적선택법(stepwise regression)

: 후진제거+전진선택

(후진제거를 하면서 이미 빠진 변수를 다시 넣을 수도 있고,
전진선택을 하면서 기존에 포함된 변수를 제거할 수도 있음)

변수선택법

- 변수선택법의 기준

1) AIC, BIC

$$AIC = \frac{1}{n\hat{\sigma}^2}(\boxed{RSS} + 2d\hat{\sigma}^2) \qquad BIC = \frac{1}{n\hat{\sigma}^2}(\boxed{RSS} + log(n)d\hat{\sigma}^2)$$

모델의 에러(RSS)가 포함된 지표(AIC, BIC가 작을수록 좋음)
설명변수의 개수가 증가하면 AIC(BIC)의 값이 커지는 패널티를 준다.
설명변수를 추가했는데 AIC, BIC값이 증가하지 않으면 좋은 변수(변수의 설명력>패널티)

2) R-square

cf) Mallow-Cp, t-test, F검정

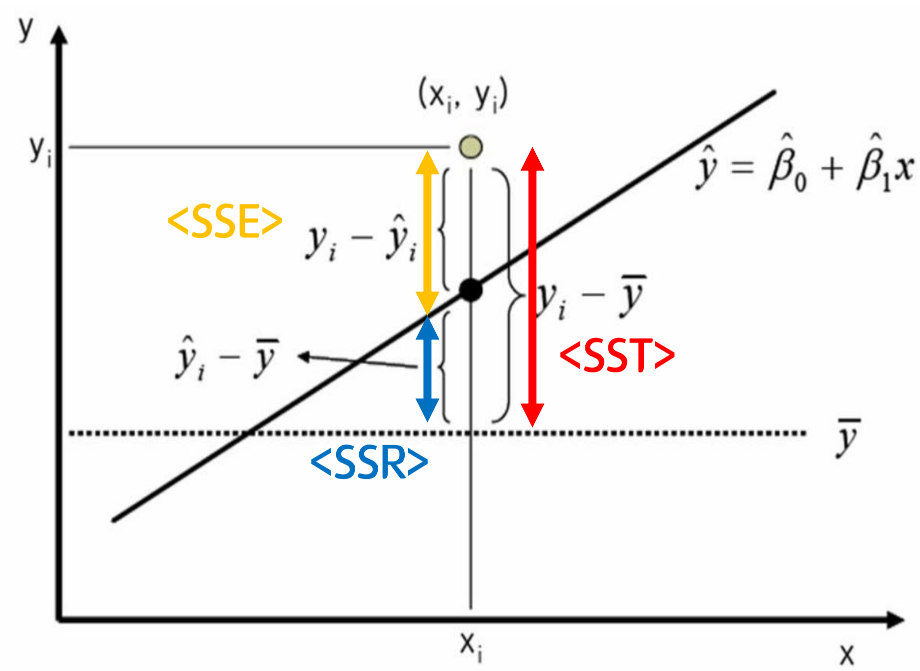
모델 적합도

- 제곱합 분해

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

<SST> <SSR> <SSE>

SST: 총제곱합
SSR: 회귀제곱합 (전체 제곱합 중 회귀식으로 설명할 수 있는 부분)
SSE: 잔차제곱합 (전체 제곱합 중 회귀식으로 설명하지 못하는 부분)



>>모델이 데이터를 잘 설명할수록 SSR증가(=SSE 감소)

모델 적합도

- R-square(=결정계수, 설명력)

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

$$\text{Adjusted } R^2 = 1 - \frac{SS_{residuals}/(n-K)}{SS_{total}/(n-1)}$$

1) R-square

전체 제곱합중 회귀식으로 설명가능한 부분
모델이 좋을수록(데이터를 잘 설명할수록) 결정계수 증가
>>결정계수가 클수록 좋음

하지만 설명변수를 추가하면 SSR이 항상 커져, 결정계수가 항상 증가
따라서 설명변수의 개수가 다른 모델의 결정계수 단순 비교불가

2)Ajdusted R-square(조정된 결정계수)

설명변수의 개수를 고려하는 R-square
설명변수가 증가하면 값이 감소하도록 패널티를 줌
>>설명변수를 추가했는데 adjusted R-square가 감소하지 않는다면, 패널티를 감수할만큼 설명을 잘하는 변수

모델 적합도

- MSE(mean squared error)

	자유도 (Degree of Freedom)	제곱합 (SS, Sum of Square)	제곱평균 (MS, Mean of Square)	F-통계량 (F-Value)
회귀 (Regression) <SSR>	k 설명변수 k개	회귀제곱합 (SSR) $\sum(\hat{y} - \overline{y})^2$	회귀제곱평균 (MSR) SSR / k	F비 MSR / MSE
잔차 (Error) <SSE>	n - k - 1	잔차제곱합 (SSE) $\sum(y - \hat{y})^2$	잔차제곱평균 (MSE) $\boxed{\text{SSE / (n - k - 1)}}$	
총 (Total) <SST>	n - 1	총제곱합 (SST) SSR + SSE	총제곱평균 (MST) SST / (n - 1)	

SSE(잔차제곱합)를 그 자유도로 나눈 값.
단순하게 생각하면 평균적인 error이다.

R-square는 회귀식이 설명하는 부분이고,
MSE는 회귀식이 설명하지 못하는 부분!

>>MSE가 작을수록 좋음

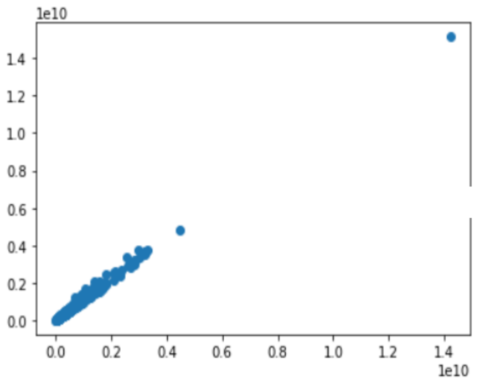
다중공선성(Multicollinearity)

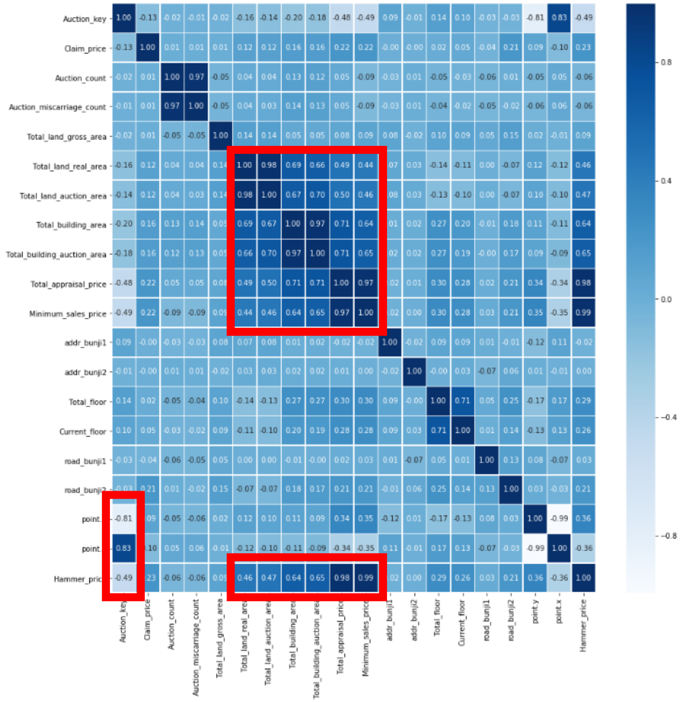
독립변수간 상관관계가 강해 독립변수의 일부를 다른 독립변수의 조합으로 표현가능한 것.

상관계수, scatter plot, VIF 등으로 확인

ex) $X_2 = 2X_1, \quad X_3 = X_1 + X_2$

다중공선성 제거 방법: 변수 삭제(선택), PCA, 정규화 등





- **VIF**(variance inflation factor)

: i번째 독립변수를 다른 독립변수들로 회귀한 성능

다른 독립변수들과 상관관계가 강할수록 VIF값이 큼

>>VIF값이 큰 변수 제거(10 이상이면 다중공선성 의심)

$$\mathrm{VIF}_i = \frac{\sigma^2}{(n-1)\mathrm{Var}[X_i]} \cdot \frac{1}{1-R_i^2}$$

Cf)회귀분석의 가정

1) 선형성

: X와 Y가 베타에 대한 선형관계이다.

2) 독립성

: 설명변수끼리 독립이다. (다중공선성이 강하면 독립성 위배)

3) 등분산성

: 오차항이 동일한 분산을 가진다.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \boxed{\epsilon_i}$$

$$Var(\epsilon_i) = \sigma^2$$

4) 정규성

: 오차항이 정규분포를 따른다. $\varepsilon \sim \mathrm{N}(0, \sigma^2)$

회귀분석 심화

다중공선성을 제거하는 이유??

- 설명변수간 독립적이지 않으면 회귀계수의 추정값이 존재하지 않을 수도 있고,
  또는 회귀계수의 추정값이 매우매우 커짐! (불안정한 추정)

$$\hat{\beta} = \boxed{(X'X)}^{-1}X'Y.$$

설명변수끼리 완벽한 선형관계가 존재하면
이부분이 Full rank가 아니여서 역행렬 존재x
완벽한 선형관계가 아니더라도 강한 다중공선성이 존재하면
이 부분이 작아서 역행렬을 취하면 값이 매우 커짐

>>따라서 회귀계수를 안정화 시켜야함

Regularization(정규화)

1) 정규화

단순한 모델을 사용하거나, 복잡한 모델을 현재 train data에 둔감하게 만들어(regularization) overfitting을 피할 수 있음. 정규화를 하면 현재 train 셋의 accuracy는 감소하지만, 잠재적인으로 test 셋의 accuracy를 증가시킴 (Variance 감소, Bias 증가)

2) Ridge regularization(L2 regularization)

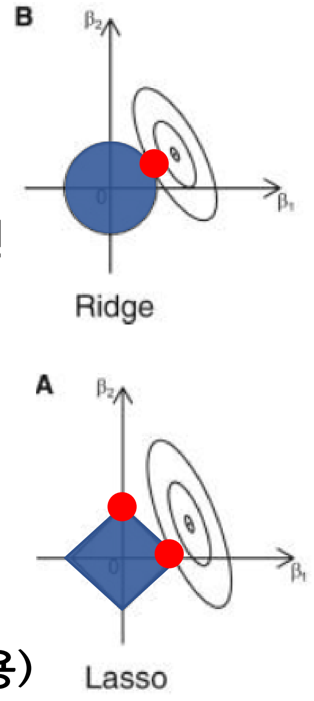
$$E(w) = \frac{1}{2}\sum_{n=0}^{N}(train_n - g(x_n, w))^2 + \underline{\frac{\lambda}{2}\|w\|^2}$$

3) Lasso regularization(L1 regularization)

$$E(w) = \frac{1}{2}\sum_{n=0}^{N}(train_n - g(x_n, w))^2 + \underline{\lambda|w|}$$

$$\frac{d}{dw}E(w) = 0$$

기존 Loss 식 뒤에 회귀계수 크기에 대한 제약조건 term이 붙은 새로운 Loss
새로운 Loss를 최소화시킴(최소제곱법)
W는 $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix},$ (회귀계수 열벡터)



(Lasso 방법은 좌표축의 한 변수만 선택되어 변수선택 방법으로도 사용)

Regularization(정규화)

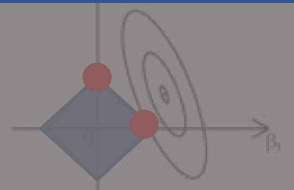
<Ridge, Lasso>

1) 회귀 계수의 크기에 제약을 줘서 회귀계수를 안정화
2) Regularization으로 overfitting을 방지

3) Lasso regularization(L1 regularization)

$$E(w) = \frac{1}{2}\sum_{n=0}^{N}(train_n - g(x_n, w))^2 + \underline{\lambda|w|}$$

Lasso 방법은 좌표축의 한 변수만 선택되어 변수선택에도 사용됨



Lasso

과제

<과제1>

- 실습 코드에서 과제 부분을 채우고, 데이터에 추가해서 모델돌리기

<과제2>

- 추정된 회귀계수를 구하는 함수 구현하기 (행렬곱을 이용해서 간단하게, numpy사용, 5줄 이하)
  input: data_x(독립변수 행렬 데이터), data_y(종속변수 벡터 데이터)
  output: 추정된 회귀계수(벡터)

<과제3>

1주차 Auction master 데이터로 회귀분석 (아래 목록들을 포함해야함)

- 자유롭게 EDA, 전처리 (저번주 과제 참고o)
- 변수 제거, 선택 시 이유 설명
- 다중공선성 확인, 처리
- fit된 모델의 평가(R-square, MSE 등등)
  (선택: 정규화, 변수선택법 등등)

Q & A

들어주셔서 감사합니다.

Reference

투빅스 11기 정규세션 회귀분석(박규리님)

투빅스 3기 정규세션 회귀분석(김상진님)

회귀분석 강의(유규상 교수님)

인공지능 및 기계학습 개론(문일철 교수님)

http://solarisailab.com/archives/1785

https://circle.haus/t/chap-4-2/103

https://datascienceschool.net/view-notebook/266d699d748847b3a3aa7b9805b846ae/

https://m.blog.naver.com/PostView.nhn?blogId=mykepzzang&logNo=220838509912&proxyReferer=https%3A%2F%2Fwww.google.com%2F

https://wikidocs.net/21670

https://datascienceschool.net/view-notebook/e6ef730b7a3b4be7be4ff028d39d67f7/

https://ko.wikipedia.org/wiki/%EC%A0%9C%EA%B3%B1%ED%95%A9

https://kmrho1103.tistory.com/entry/%EC%A0%9C2%EC%9E%A5-%EC%A4%91%ED%9A%8C%EA%B7%80%EB%AA%A8%ED%98%95-%EC%A4%91%ED%9A%8C%EA%B7%80%EB%AA%A8%ED%98%95

https://cinema4dr12.tistory.com/1275