


Linear Regression

1. Classification and Regression
2. Linear Regression - 특징, 도출방법(MSE), Cost and optimizer
3. feature selection, (coefficient)
4. Fitting accuracy, p-value, adjusted R², MSE

1. 지도학습 Unsupervised

답을 알려주고 학습 target value가 없어

1. 주가 1. 비슷한 사진 찾기

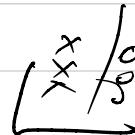
2. 사진 2. 고객데이터 행동 패턴 - associations

3. 사용族群 Grouping - Clustering

4. no target data

Classification - Category

등락



지도

Regression - real value

연속형

증가율



선형 vs 비선형



2. Linear Regression

ex)



$$y = f(x)$$

절편

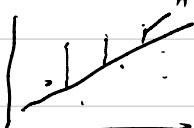
$$y = \text{연차} \times \text{가장 치} \times \text{연차} + b$$

$$y = w_1 x_1 + w_2 x_2 + w_3 x_3 + b - \text{이제 선형}$$

그리고 w 를 찾고 가장 설명을 잘하는 직선 찾기

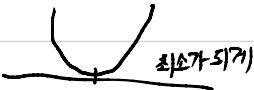
거리 차이로 선을 찾아 잔차 residual

잔차의 제곱의 최소



Gradient descent

$$\text{cost} \quad \text{MSE} = \frac{1}{n} \sum (y - \hat{y})^2$$



$$\text{Chain rule!} \quad \frac{\partial \text{cost}}{\partial z} = \frac{\partial f(z)}{\partial z} \cdot \frac{\partial z}{\partial w} \quad \frac{\partial f}{\partial z} \frac{\partial z}{\partial w} = \frac{1}{n} \sum (wz^i + b - y^i) [wz^i + b - y^i]'$$

$$(f \circ g)'(x) = f'(g(x)) \times g'(x)$$

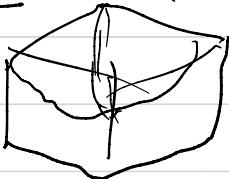
$$= \frac{1}{n} \sum (w \cdot x^{(i)} + b - y^{(i)}) x^{(i)}$$

$$w = w - \alpha \frac{\partial}{\partial w} \text{cost}(w)$$

$$w = w - \alpha \frac{1}{m} \sum_{i=1}^m (wz^{(i)} + b - y^{(i)}) x^{(i)}$$

$$w_i : w_{i-1} \xrightarrow{\text{正則}} \frac{\partial}{\partial w} \cos f(w)$$

more optimizer
in deep learning

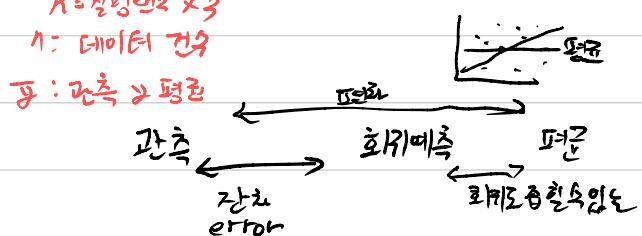


자유도	제곱합	제곱평균	F
회귀	df	$SSR = \sum (\hat{y} - y)^2$	MSR
잔차	$n - df - 1$	$SEE = \sqrt{\sum (y - \hat{y})^2}$	MSE
총	$n - 1$	$SST = SEE + SSR$	MST

$$K = \text{설명변수} \times 3$$

7. 데이터 구조

平：古字又譯卷



kernel  +  why
설명 강회기도 고차원입니다.

MSE: 회귀로 설명x

F : 크면 좋아

MSR: 최적설명 가능 편도

구글하나 Regression에서

accuracy score 는
MSE 외에 있으면 OK인가?

3 평가

$$R^2 = \frac{SSR}{SST} = \frac{SST - SSE}{SST} \rightarrow \text{이에 아까운 수록 좋아}$$

$$\text{adj } R^2 = 1 - \frac{MSE}{MST}$$

whg 도입 변수들 빼면 R² 높아지

그에서 자유도나는면서 $\text{adj } R^2 \geq R^2$

· 다중공선성 - 독립변수간 높은 달관관계 - 상관관계와 분석방법 접근

문제점: 회귀계수 불안정 해결책- 피어슨 상관계수 Matrix로 OLS

서리성 안정성

VIF > 10

· Feature selection은 알아서;

more data
feature reduce

4. Regulation - L1, L2 to avoid overfitting early stopping, dropout

$$\text{Norm } \|x\|_p := \left(\sum_{i=1}^n |x_i|^p \right)^{1/p} \quad p=1 \text{ L1} \quad p=2 \text{ L2}$$

L₁ Norm $d_1(p, q) = \sum_{i=1}^n |p_i - q_i|$ 차이의 절대값

$$\|x\|_2 = \sqrt{x_1^2 + \dots + x_n^2} \Rightarrow \text{직교기저}$$



L1 Loss

$$\sum_{i=1}^n |y_i - f(x_i)|$$

L2

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

Lasso

제한된

cost function에 대해서 가중치가 너무 커지지 않게,

$$Ridge = \alpha \sum_{j=1}^n |\theta_j|$$

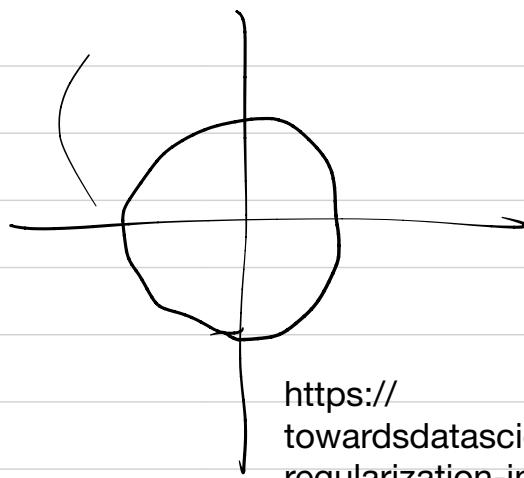
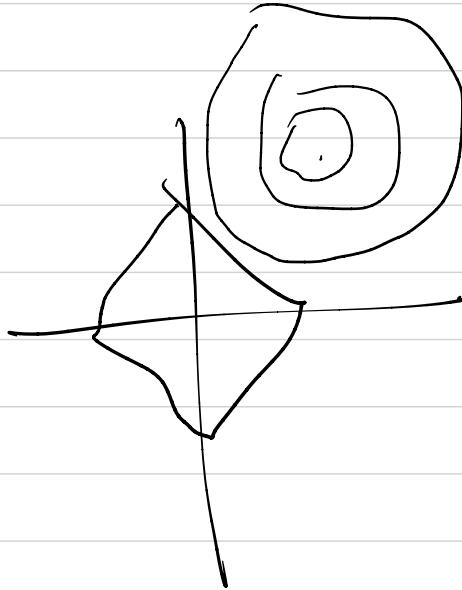
gradient Des L1 regul

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m (h_\theta(x^i) - y^i)^2 + \lambda \sum_{j=1}^n |\theta_j|$$

$$\theta^* = \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^m (h_\theta(x^i) - y^i) x^i - \gamma$$

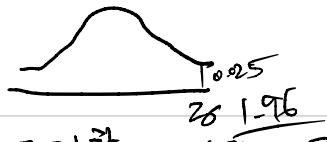
L2

$$-2\gamma\theta$$



<https://towardsdatascience.com/regularization-in-machine-learning-connecting-the-dots-c6e030bfadd>

Pvalue



기후가설이 맞다는 전제하, 같은 대상군간 통계치 관찰확률
- 통계적 통증과 같은 일정[2]