

05 회귀

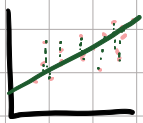
01. 회귀 소개

회귀 (Regression)

: 여러개의 독립변수나 한개의 종속변수 간의 상관관계를 모델링하는 기법. → "회귀 회계계수" 찾기

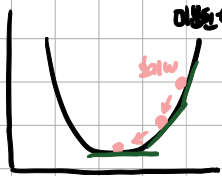
- 단변 / 다변. 선형 / 비선형

02. 다변선형회귀를 통한 회귀 이해



$$Error^2 = RSS$$

03. 비용 최소화하기 - 경사하강법 (Gradient Descent)



비용함수의 기울기가
최소인 점

- 새로운 $w_1 =$ 이전 $w_1 +$ 학습률 $\cdot \frac{2}{N} \sum_{i=1}^N x_i (y_i - \hat{y}_i)$

- 새로운 $w_0 =$ 이전 $w_0 +$ 학습률 $\cdot \frac{2}{N} \sum_{i=1}^N (x_i y_i - \hat{y}_i)$

- 미니배치 학습을 경사하강법 → 시간↓

04. 사이킷런 LinearRegression을 이용한 보통의 주택가격 예측

• 회귀 평가 지표

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

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

$$RMSE = \sqrt{MSE} + RMSLE$$

$$R^2 = \frac{\text{예측값 Var}}{\text{실제값 Var}}$$

• 다항식회수 ~ [복잡한 피상관 관계 모델링 가능
과제형의 문제 많음

• 편향 - 분산 트레이드 오프
→ 오류 cost 최대로 낮추는 모델 찾기!

06. 규제선형모델 - 릿지, 라소, 엘라스틱넷

• 학습데이터 과소적합 최소화
회계계수 크기 제한 > 사이 균형 필요

⇒ 과적합 위해 $RSS(w) + \alpha ||w||^2$ 최소화

- $\alpha=0 \rightarrow$ 0 = 0

$\alpha=\infty \rightarrow$ 0 = $\infty \rightarrow w$ 들이 0가깝게

- L2규제: w 의 제곱에 페널티 부여
= "릿지 회귀"

• L1규제: $\alpha \cdot ||w||_1$. w 의 절대값에 페널티 부여
= "라소 회귀"

- 필요한 회계계수 급격 감소 → 0
적당한 도너만 회계계수 포함

• "엘라스틱넷 회귀" = L1규제 + L2규제

• 선형선형모델을 위한 데이터 변환

- StandardScaler, Min/MaxScaler: 성능향상

- 스케일링, 정규화 속성 → L1/L2특성 적용: 과적합 예방

- 로그변환: 가파른 기울기 완화

07. 크리스티 트리

- 분류에 사용
- 시그모이드 함수 확률 \rightarrow 반비례 함수 확률로
- 장점
 - 가볍고 빠름
 - 아전 분류 예측 성능 good. \rightarrow 아전 분류기보다
 - 호전 데이터 분류 성능 good \rightarrow 텍스트 분류

08. 회귀 트리

- 분류 트리보다 빠름
- but, 리프 노드 N 데이터의 평균값을 회귀 예측값 계산

