14. Surrogate Models

- 전통적인 최적화 : 함수 f을 알고 있음
- Surrogate model : 목적 함수 f 모음
- m개의 계획점과 이에 대응하는 함수의 값 y
- 목적 함수 : $\underset{\theta}{\text{minimize}} \; \| y - \hat{y} \|_p$

L_p norm

- Surrogate Models

- m개의 계획점

  : $X = \{x^{(1)}, x^{(2)}, \cdots, x^{(m)}\}$

- m개의 함수값

  : $y = \{y^{(1)}, y^{(2)}, \cdots, y^{(m)}\}$

- m개의 예측값 : $\hat{f}$, θ (모수)

  $\hat{y} = \{\hat{f}_\theta(x^{(1)}), \hat{f}_\theta(x^{(2)}), \cdots, \hat{f}_\theta(x^{(m)})\}$

⇒최적화 문제는 계획점 찾기 아니고

함수 f라고 생각되는 surrogate 함수 구성하는

θ 모수 찾기

- 목적 함수 (p=2 : 유클리디안)

  : $\underset{\theta}{\text{minimize}} \; \|y - \hat{y}\|_p$, L_p norm

- 선형 모형 (linear models)

  $\hat{f} = w_0 + w^T x$, $\theta = \{w_0, w\}$ ← 모수!

  $\hat{f} = \theta^T x$

- 선형회귀 (Linear regression)

  : $\underset{\theta}{\text{minimize}} \; \|y - \hat{y}\|_2^2$ → 최소 제곱 합

- 행렬 표현

  : $\underset{\theta}{\text{minimize}} \; \|y - X\theta\|_2^2$

  → 계획 행렬 $X = \begin{bmatrix} (x^{(1)})^T \\ \vdots \\ (x^{(m)})^T \end{bmatrix}$ 계획점 행렬

- Analytic Solution : 최소 제곱 추정량 (Newton 방법의 특수한 경우)

  $\underset{\theta}{\text{minimize}} \; \|y - X\theta\|_2^2$

  → $\theta = (X^T X)^{-1} X^T y$ ↑ 최소제곱 추정량 구하기 어렵 → 역행렬!

  → 대규모 자료 : m이 매우 크고 X 차원 매우 높음

  → Newton 방법의 특수한 경우

  ⇓

  → $(X^T X)^{-1}$: 이차 미분 정보 사용X (Hessian), gradient 사용하자!

  → 경사도

  $2X^T(y - X\theta) \rightarrow (X\theta - y)X^T$

  $2X^T(y - X\theta) \rightarrow (\hat{y} - y)X^T$

  → Full-batch Gradient Descent : X와 y 이용

  $\theta^{(t+1)} = \theta^{(t)} - \alpha(\hat{y} - y)X^T$

  계획 행렬의 전체 data

- Stochastic Gradient Descent

- 한 개의 계획점 X 와 함수값 y

- 경사도 (계획점 하나 → mini batch)

  $(\hat{y}_{(i)} - y_{(i)}) \cdot x_{(i)}^T \Rightarrow (\hat{y}_{(i)} - y_{(i)}) \cdot x_{(j)}$

- SGD

  $\theta_j^{(t+1)} = \theta_j^{(t)} - \alpha(\hat{y}_{(i)}^{(t)} - y_{(i)}) \cdot x_{(j)}$

- Mini batch Stochastic Gradient Descent

- 전체 자료를 m보다 작은 임의 부분 집합들로 나눔

  ex) 100개 자료 → 10개씩 10번의 미니배치

  $\theta_j^{(k+1)} = \theta_j^{(k)} - \alpha \sum_{k=i \times B}^{(i+1) \times B} (\hat{y}_{(k)}^{(t)} - y_{(k)}) \cdot x_{(k)}$

  → adam 방법 자주 사용