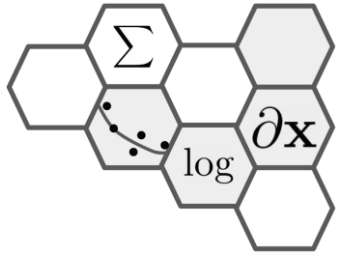


LG U+ Why Not SW 캠프 6기
Python 데이터 분석 I

Linear Regression

조준우
metamath@gmail.com

머신러닝이란?

https://medium.com/@ezralazuardy/how-machine-learn-c2f73f60ef14

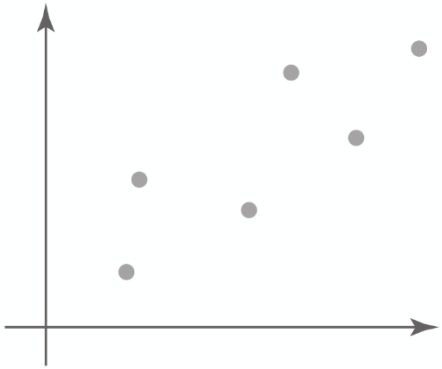
- 넓은 의미 : 컴퓨터를 이용한 문제 해결에 있어서 지식기반, 규칙 기반 방법이 아닌 데이터와 범용 알고리즘으로 퍼포먼스를 개선하는 방법

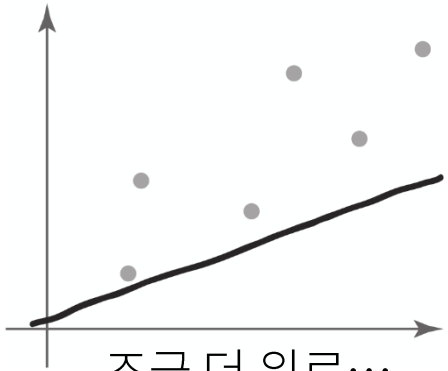
- Samuel, A. L. (1959) : "Programming computers to learn from experience should eventually eliminate the need for much of this detailed programming effort"
  “Some Studies in Machine Learning Using the Game of Checkers” in IBM Journal of Research and Development (Volume:3, Issue: 3), p. 210, 기계학습, 오일석, 한빛미디어

- Tom Mitchell (1998) : Well-posed Learning Problem: A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.
  Mitchell, T., 1997, Machine Learning, McGraw Hill

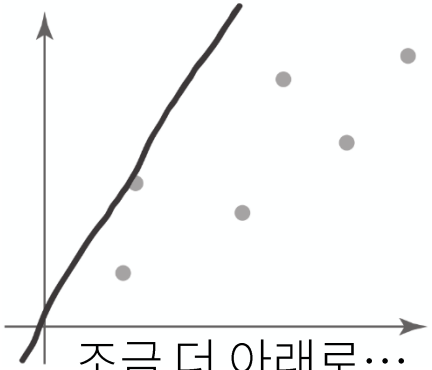
- 주어진 데이터를 가장 잘 표현하는 함수를 찾는 것

회귀 문제: 아이가 선을 그린다면…

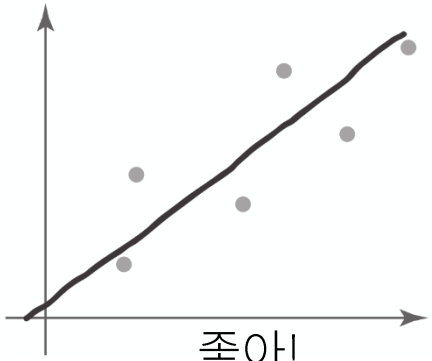




조금 더 위로…



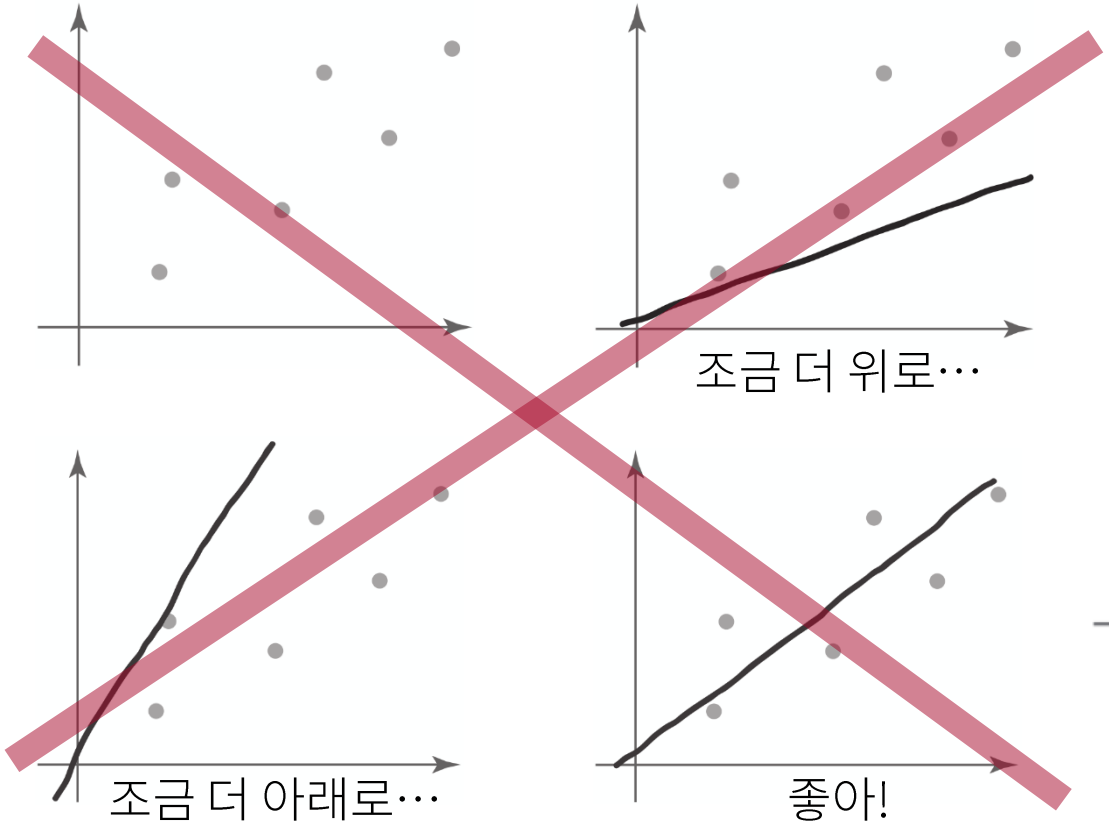
조금 더 아래로…

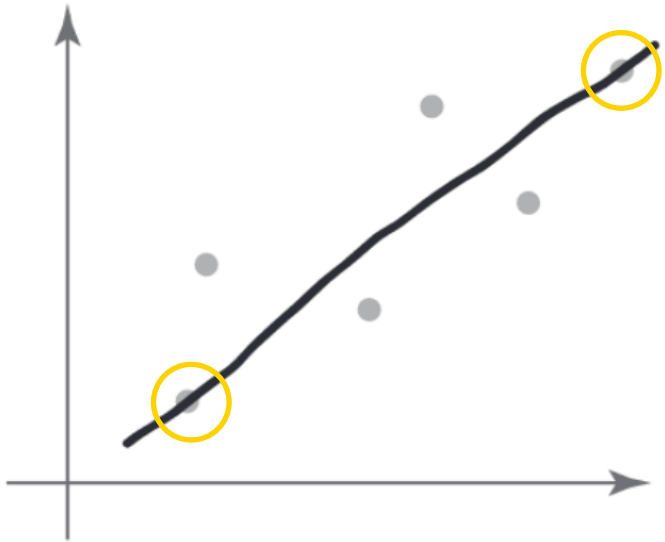


좋아!

- Task T : 선을 그린다.
- Measure P : 선이 점에 잘 맞는가?
- Experience E : 만족할 때 까지 계속 선을 그린다. 점에서 선이 멀어지면 좋지 않다.

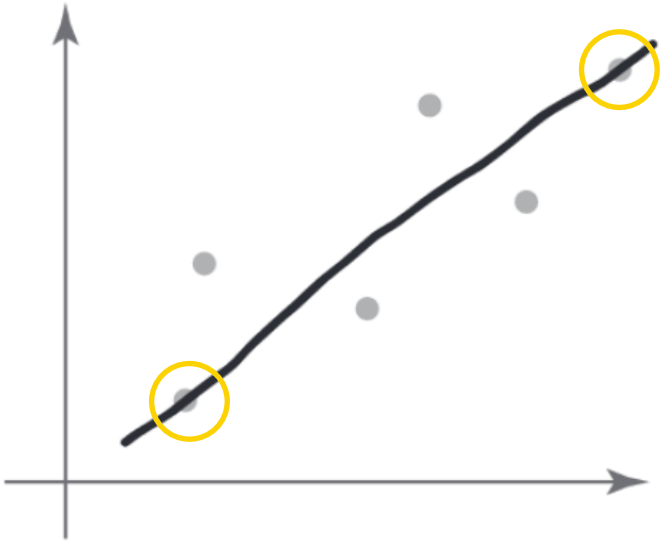
조금 더 똑똑하게?: 규칙기반 알고리즘



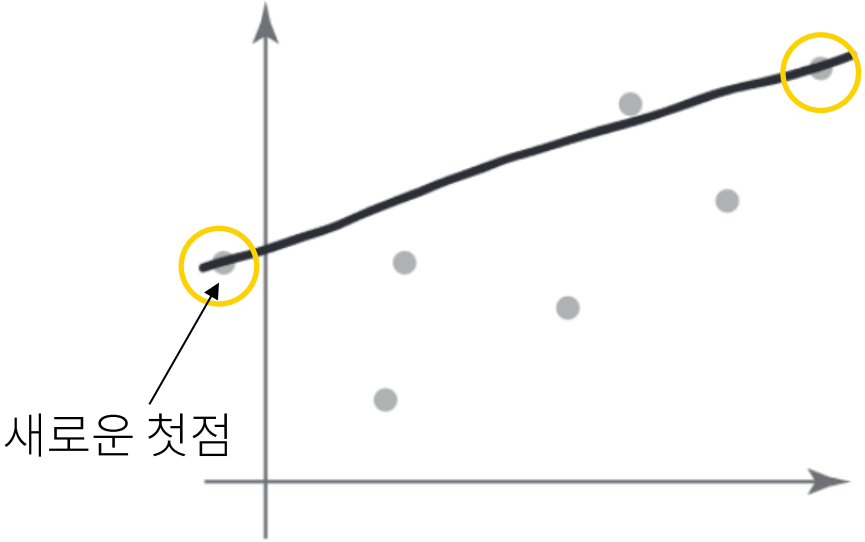


지루한 반복없이 첫점과 끝점 연결

정말 똑똑한가?

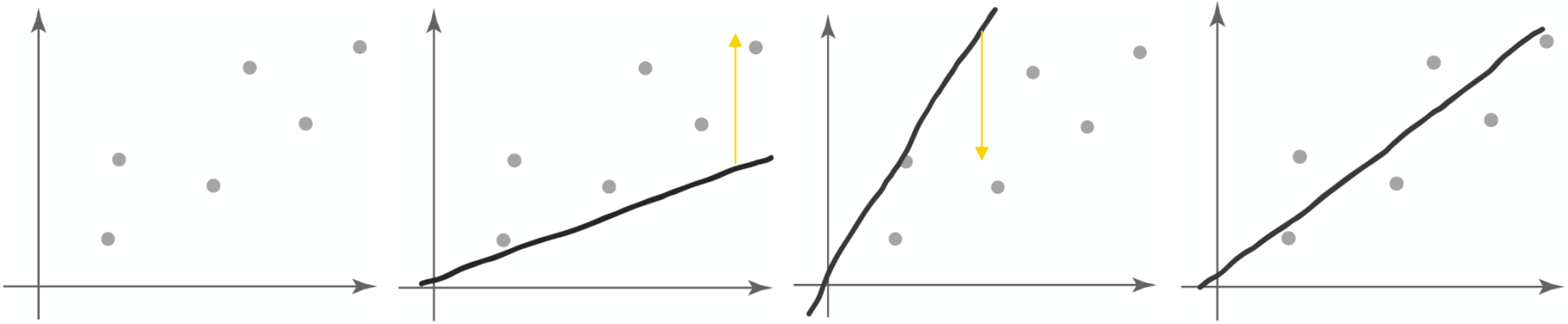


지루한 반복없이 좋아!



음…?
혹시 머신러닝으로는?

선형회귀 맛보기



직선 표현

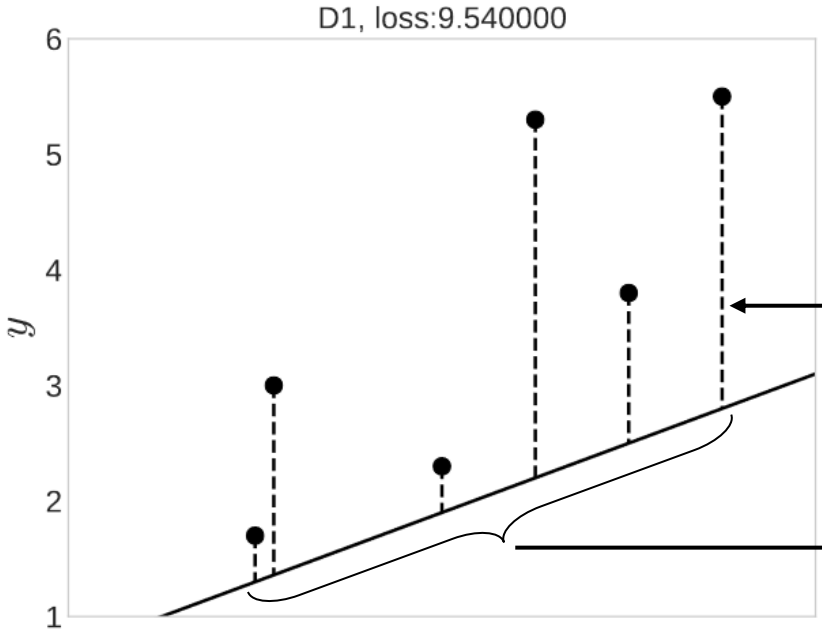
$$h(x, \mathbf{w}) = w_1 x + w_0$$

직선 움직임

$$h(x, \mathbf{w}) = w_1 x + w_0$$

w_1 w_0

얼마나 안 좋은가?

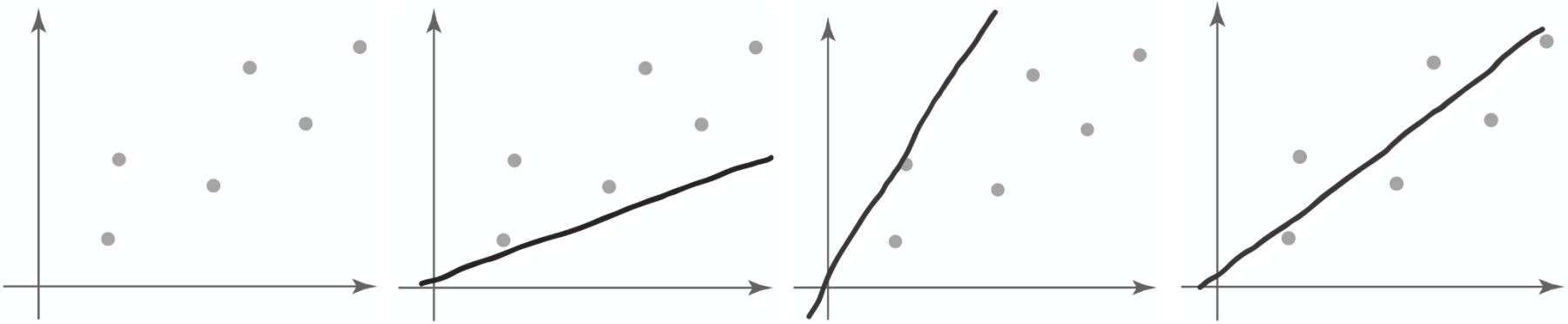


$$L = \frac{1}{2}\sum_{n=1}^{N}\{h(x_n, w) - y_n\}^2$$

점과 선의 수직거리

모든 점에 대해서 다 더함

다이얼을 돌리자!



다이얼을 돌림
Update

$$\mathbf{w} \leftarrow \mathbf{w} - \eta \frac{\partial}{\partial \mathbf{w}} \left(\frac{1}{2} \sum_{n=1}^{N} \{h(x_n, \mathbf{w}) - y_n\}^2 \right)$$

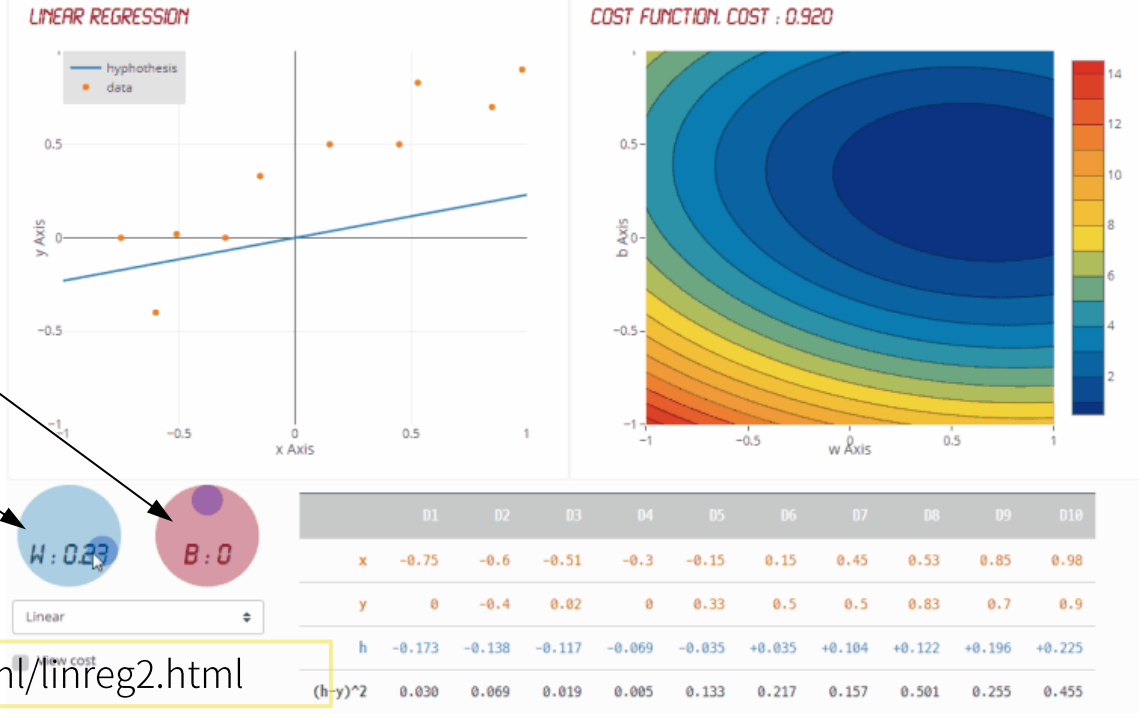
학습률Learning
Rate

경사도
Gradient

손으로 선형회귀…

- 다이얼을 돌리면서 가장 알맞은 직선 찾기

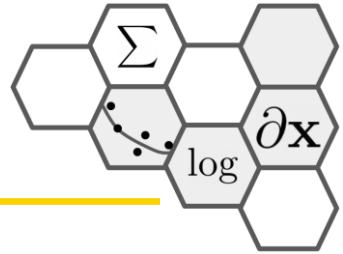
$$h(x, \mathbf{w}) = w_1 x + w_0$$



	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
x	-0.75	-0.6	-0.51	-0.3	-0.15	0.15	0.45	0.53	0.85	0.98
y	0	-0.4	0.02	0	0.33	0.5	0.5	0.83	0.7	0.9
h	-0.173	-0.138	-0.117	-0.069	-0.035	+0.035	+0.104	+0.122	+0.196	+0.225
(h-y)^2	0.030	0.069	0.019	0.005	0.133	0.217	0.157	0.501	0.255	0.455

JS metamath1.github.io/noviceml/linreg2.html

머신러닝 분류: 지도학습



- 지도학습Supervised learning
  - 회귀 문제Regression : 선형 회귀Linear Regression
    - 정답 : 연속된 실수
    - 예 : 대지면적에 따른 집값, 시간에 따른 트랜지스터 집적 개수, 기온에 따른 빙과류 판매량
  - 분류 문제Classification : 로지스틱 회귀Logistic Regression
    - 정답 : (0,1) 또는 (0, 1, …, K)
    - 예 : 개-고양이 분류, 양성종양-악성종양 분류, …