

# SGDClassifier

확률적경사하강법

# 확률적 경사하강법(주요 파라미터)

```
SGDClassifier(alpha, average, class_weight, epsilon, eta0, fit_intercept, l1_ratio,
learning_rate,
loss, max_iter, n_iter, n_jobs, penalty, power_t, random_state, shuffle, tol,
verbose, warm_start)
```

- loss : 손실함수 (default='hinge' -> SVM에서 주로 사용)
- penalty : { 'l2' , 'l1' , 'elasticnet' }, default= 'l2'
- alpha : 값이 클수록 강력한 정규화(규제) 설정 (default=0.0001)
- max\_iter : 계산에 사용할 작업 수 (default=1000)
- fit\_intercept : 모형에 상수항 (절편)이 있는가 없는가를 결정하는 인수 (default=True)
- learning\_rate : 학습률 (default = 'optimal')

# 파라미터 간단한 설명

Regulation을 하는 이유 : 과대적합을 방지하고 일반성을 띄게 해주기 위해

- L1penalty : Lasso

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

Cost function에 가중치의 절대값을 더함

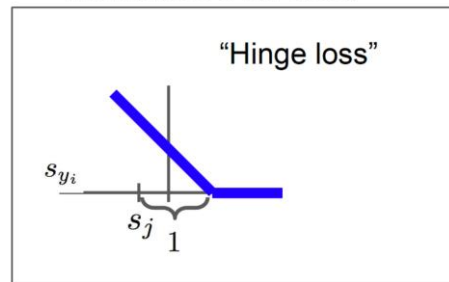
- L2penalty : Ridge

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

Cost function에 가중치의 제곱값을 더함

- Hinge Loss

Multiclass SVM loss:



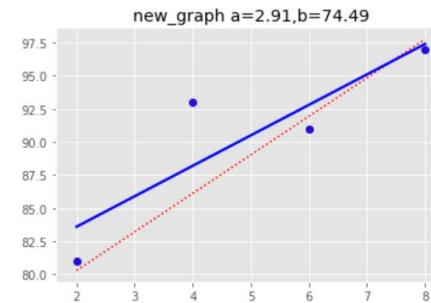
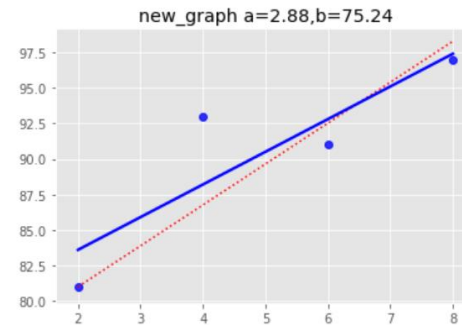
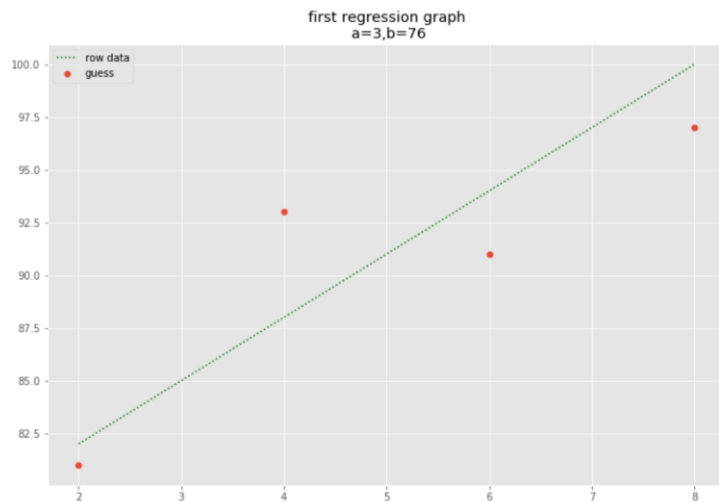
X축은 정답 라벨의 점수  
Y축은 loss

$$L_i = \sum_{j \neq y_i} \begin{cases} 0 & \text{if } s_{y_i} \geq s_j + 1 \\ s_j - s_{y_i} + 1 & \text{otherwise} \end{cases}$$
$$= \sum_{j \neq y_i} \max(0, \underline{s_j - s_{y_i} + 1})$$

엄격하게 판단하기 위해 1을 더함

# What is Gradient Descent

- 함수 값이 낮아지는 방향으로 독립 변수 값을 변형시켜 가면서 최종적으로는 최소 함수 값을 갖도록 하는 독립 변수 값을 찾는 방법

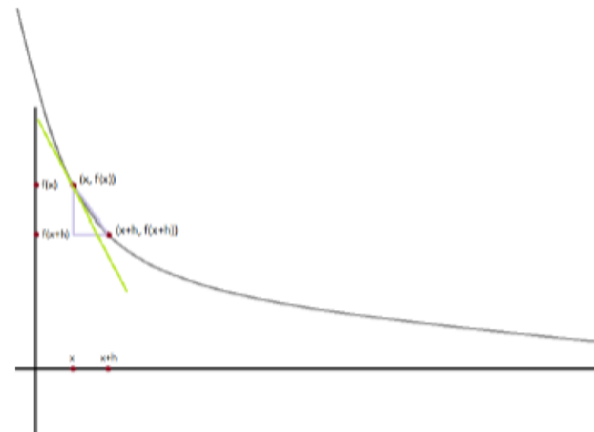


# Estimating the Gradient

미분 값을 조절하며 접하지 않았던 선을 점점 함수와 접하게 하면서 최적의 기울기를 추출한다.

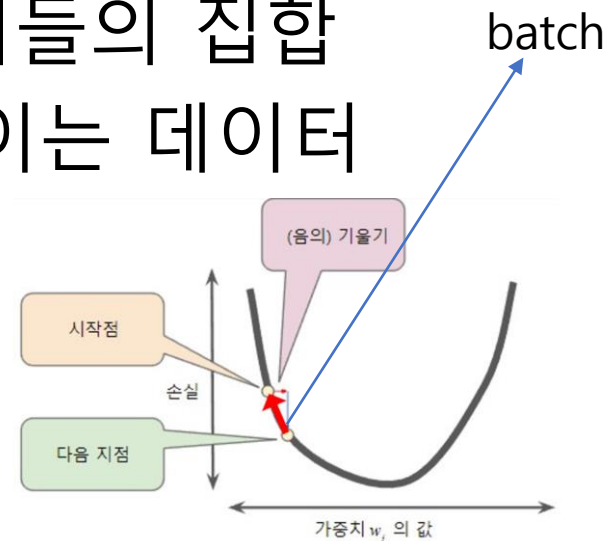
임의의 시작점을 잡고 Gradient가 아주 작아질 때까지 경사의 반대 방향으로 조금씩 이동하면 벡터의 최솟값을 추출할 수 있음

$$x_{i+1} = x_i - \alpha \frac{df}{dx}(x_i)$$



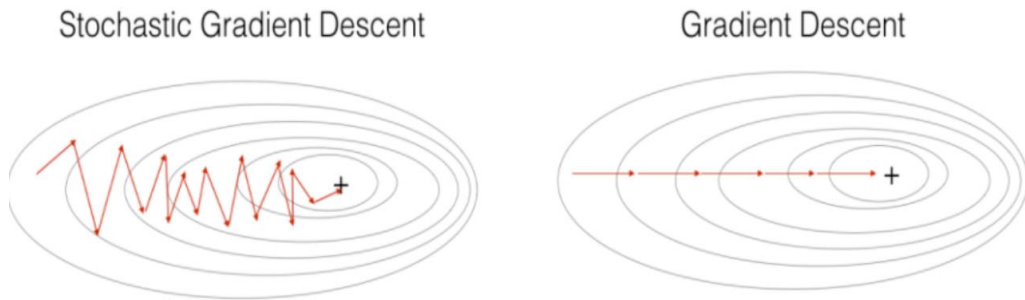
# Stochastic Gradient Descent

- 용어 개념 정리
- Batch : 경사 업데이트 1회에 사용되는 예시들의 집합  
-> 딱 한 번 다음 점으로 이동하는데 쓰이는 데이터
- 만약 배치가 크다면 Random Sampling한 데이터 세트엔 중복 데이터가 포함 가능성이 크다.  
따라서 배치 크기를 조절하여 학습!!!



# Stochastic Gradient Descent 특징

- Gradient를 구하는데 1개의 데이터만 사용(batch\_size = 1)
- 데이터는 무작위로 선택(따라서 확률적)
- 반복이 충분할 경우 효과 있음
- 노이즈가 심함
- 각 반복마다 무작위로 선택

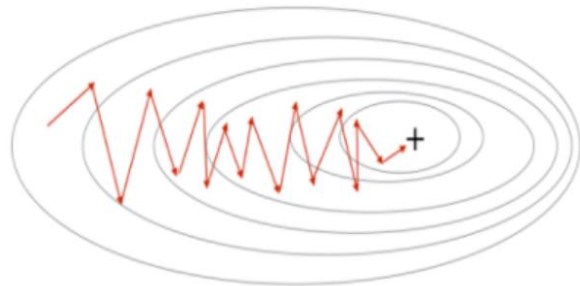


노이즈가 많고 여러 번의 반복 과정이 필요하지만  
Batch\_size=1이므로 **계산이 더 빠르다**.

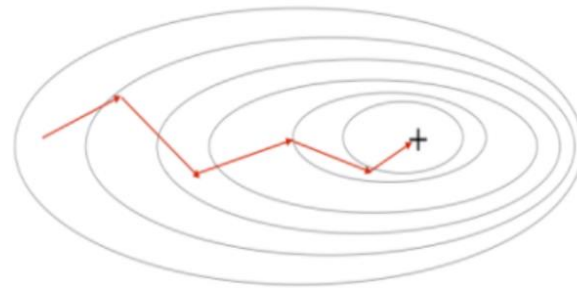
# 미니 배치 확률적 경사 하강법

- 배치 사이즈가 1이므로 너무 작아 반복량이 많아야한다.
- 따라서 이러한 문제점을 줄이기 위해 작은 배치 사이즈로 경사 하강법을 진행

Stochastic Gradient Descent



Mini-Batch Gradient Descent





# Visualization

