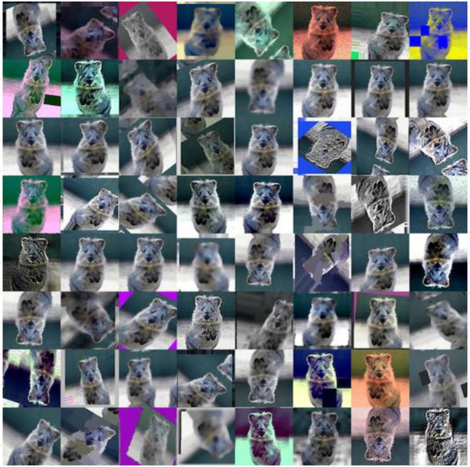
Khyeyoon

Data agumentation



출처 : https://www.kakaobrain.com/blog/64

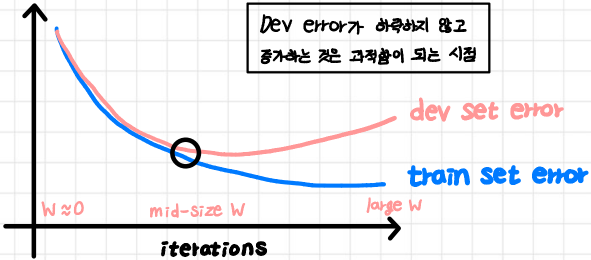
- Random Horizontal Flip
- Random vertical Flip
- Random Affine
- Random crop
- Random Resized Crop
- Random Gray Scale
- Random Rotation

⋮

정규화 기법과 비슷한 효과를 냄

Khyeyoon

Early Stopping



조기종료는 개발세트의 오차가 저점일 때 훈련을 멈추는 것

→ 가장 잘 작동하는 점에서 멈추기

Orthogonalization

- **optimize cost function J**
  - ex) gradient descent, momentum, RMSProp, Adam
- **Not overfit**
  - ex) regularization, more data

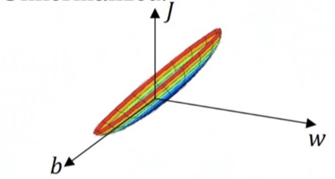
두가지 작업은 별도의 일이라서 다른 방법으로 접근해야 하지만 조기종료는 두가지 작업을 섞어버리기 때문에 최적의 조건을 찾기 어려울 수 있음

Khyeyoon

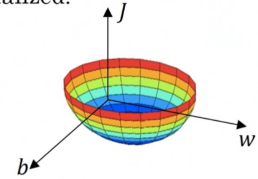
Why normalize inputs?

$$J(w,b) = \frac{1}{m}\sum_{i=1}^{m} L(\hat{y}^{(i)}, y^{(i)})$$

Unnormalized:

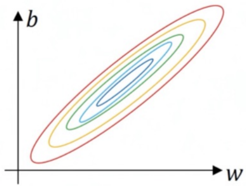


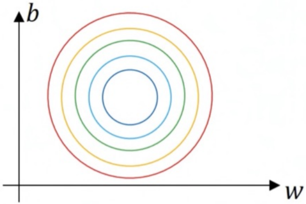
Normalized:



정규화를 통해 비용함수의 모양은 더 둥글고 최적화하기 쉬운 모습이 됨

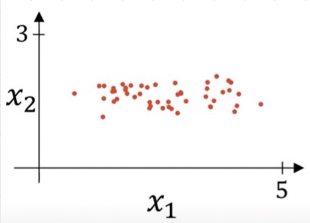
→ 효율적인 학습



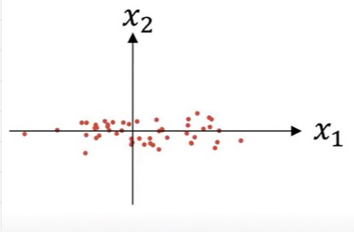


Khyeyoon

Normalizing training sets



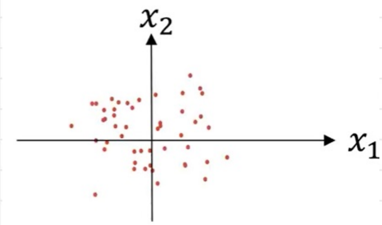
Original data



평균을 0으로 만들기

$$\mu = \frac{1}{m}\sum_{i=1}^{m} x^{(i)}$$

$$x := x - \mu$$



분산을 1로 만들기

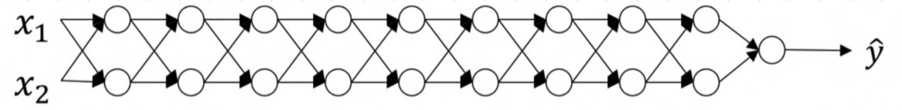
$$\sigma^2 = \frac{1}{m}\sum_{i=1}^{m} x^{(i)2}$$

$$x := \frac{x}{\sigma^2}$$

Test sets를 정규화할 때도 동일하게 정규화하기 (동일 μ σ 사용)

Khyeyoon

Vanishing / Exploding gradients



$g(z)=z$, $b^{[l]}=0$ 이라고 가정,

$$\hat{y} = W^{[l]}W^{[l-1]}W^{[l-2]}\cdots W^{[3]}W^{[2]}\underbrace{W^{[1]}X}_{z^{[1]}}$$

$$\begin{bmatrix} 1.5 & 0 \\ 0 & 1.5 \end{bmatrix}$$

모든 가중치 행렬 $\boxed{W=1.5E}$ 라고 가정하면,

$\hat{y} = 1.5^{(l-1)}Ex$

신경망이 깊어질수록 $\hat{y}$의 값은 기하급수적으로 증가

"exploding gradient"

$$\begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$$

모든 가중치 행렬 $\boxed{W=0.5E}$ 라고 가정하면,

$\hat{y} = 0.5^{(l-1)}Ex$

신경망이 깊어질수록 $\hat{y}$의 값은 기하급수적으로 감소

"Vanishing gradient"