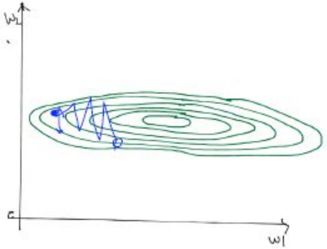
머신러닝 스터디 4th week 보조 자료

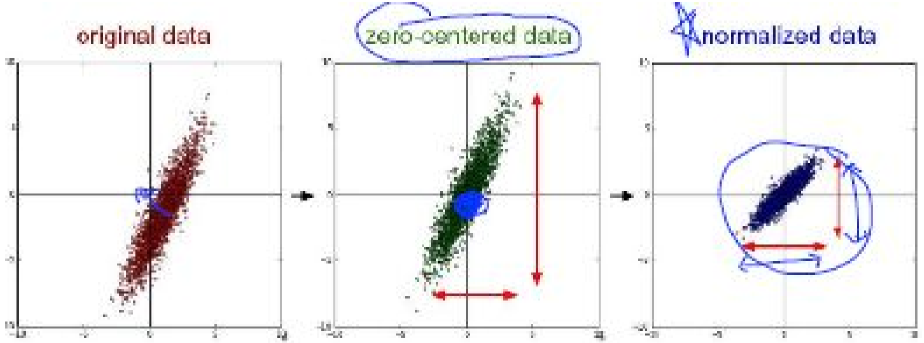
20180125 김성헌

summary

- ch.7 Application & Tip
  - Learning rate
    - cost 함수의 값을 관찰하고 합리적인 rate를 찾아라. 0.01을 사용해보고 조금씩 조절해라.
  - data preprocessing
    - 특징 간 값 차이를 줄여라
    - normalization (Standardization)

x1	x2	y
1	9000	A
2	-5000	A
4	-2000	B
6	8000	B
9	9000	C



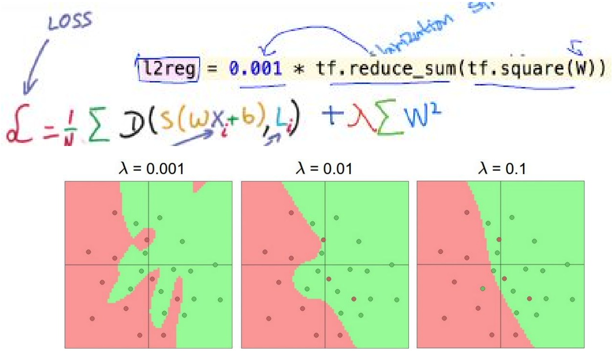


$$x'_j = \frac{x_j - \mu_j}{\sigma_j}$$

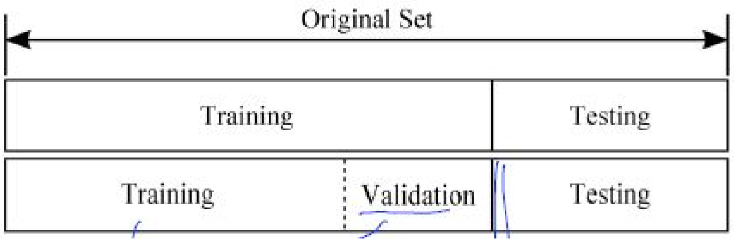
```
X_std[:,0] = (X[:,0] - X[:,0].mean()) / X[:,0].std()
```

summary

- ch.7 Application & Tip
    - overfitting
        - 더 많은 학습데이터를 사용해라
        - 특징의 수를 줄여라.
        - Regularization

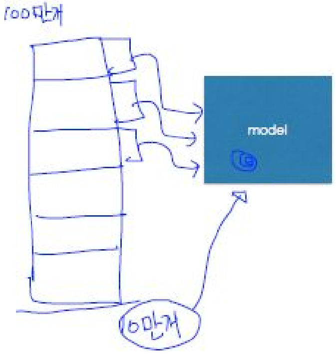


    - Learning and test data sets



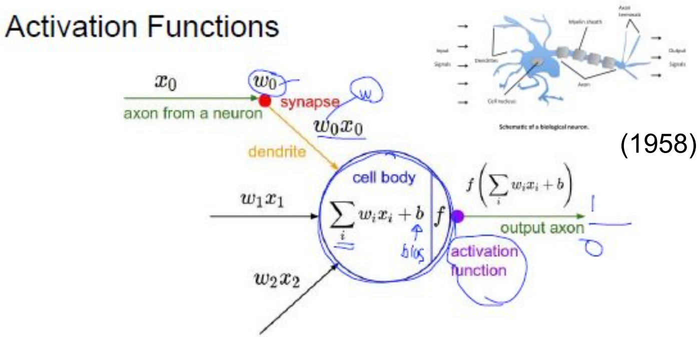
validation set으로
학습 파라메터 튜닝 해라

    - Online learning

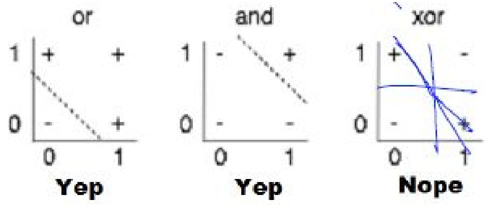


summary

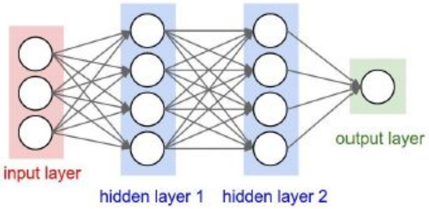
- ch.8 Neural Nets



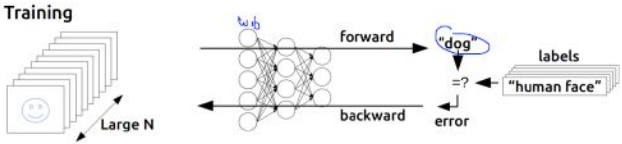
(1958)



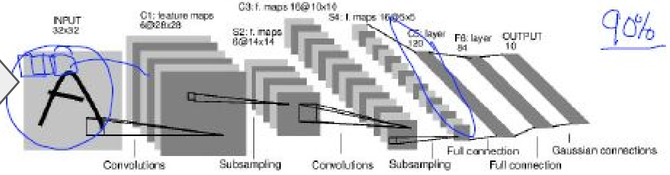
logistic regression unit 으로는
XOR를 못 푸네...



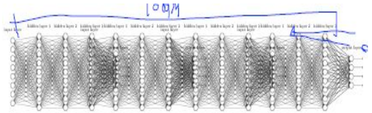
MLP를 사용하면 XOR을 풀수 있지만
학습이 어려워... Minsky(1969)



Backpropagation으로 학습 할 수 있잖아...
Hinton(1986)

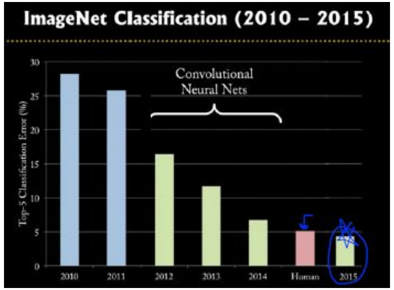


고양이 뇌처럼 필요한 뉴런만 사용해 보자.
LeCun(1980)

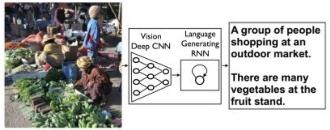


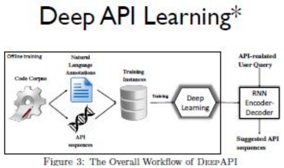
그런데 layer가 너무 많으니까
학습이 잘 안되.
대신 SVM, RandomForest
등이 간단한 알고리즘으로
성능도 좋아.

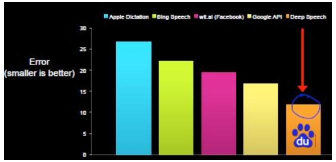
w 초기값을 잘
주면 가능해 ..
Hinton(2006)
Deep Nets.
Deep Learning
으로 부르자



Deep Nets은 성능이 정말
좋잖아 (2012 ~)







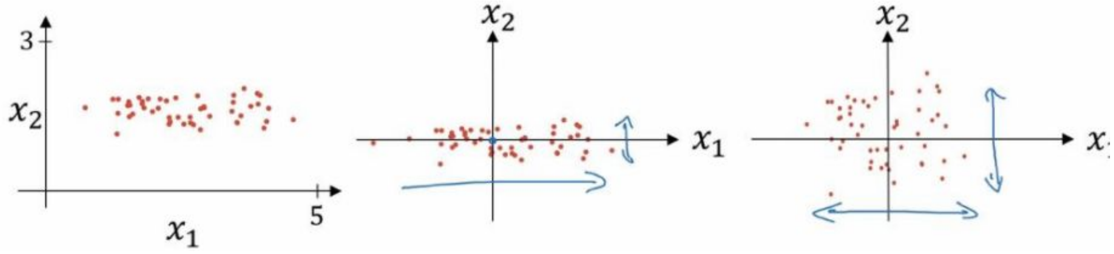
Data preprocessing 보충

- Normalization
  - 0~1 사이의 값으로 나타내는 척도법

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

- Standardization

$$x_{new} = \frac{x - \mu}{\sigma}$$

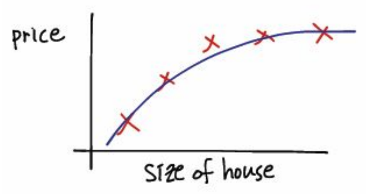


Regularization 보충

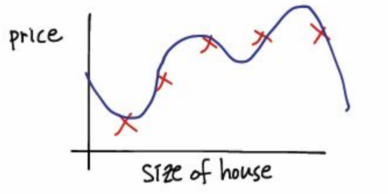
- 종류
  - L1: $R(W) = \sum_k \sum_l |W_{k,l}|$
  - L2: $R(W) = \sum_k \sum_l W_{k,l}^2$
  - Elastic: $R(W) = \sum_k \sum_l \beta W_{k,l}^2 + |W_{k,l}|$
    (L1과 L2의 혼합)
- Andrew Ng 자료
  -

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2 + \lambda \sum_{i=1}^{100} \theta_j^2$$

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



$\theta_0 + \theta_1 x + \theta_2 x^2$



$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$

Gradient Descent

Repeat {

θ_0는 regularize 안함, by convention

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y \right)^2 x_0^{(i)}$$

$$\theta_j := \theta_j - \alpha \underbrace{\left[\frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2 x_j^{(i)} + \frac{\lambda}{m} \theta_j \right]}_{\frac{\partial}{\partial \theta_j} J(\theta)}$$

$j = \cancel{0}, 1, \cdots, n$

}

θ_j 항을 모아서 쓰면,

$$\theta_j := \theta_j \underbrace{\left(1 - \alpha \frac{\lambda}{m}\right)}_{<1} - \alpha \frac{1}{m} \sum_i \left(h_\theta(x^{(i)}) - y^{(i)} \right) x_j^{(i)}$$

for $\alpha > 0$, $\lambda > 0$, $m > 0$

원래 gd 에서 하던 역할 그대로

θ_j 가 점점 작아지는 효과