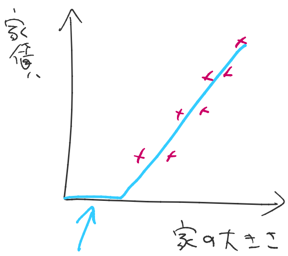
Course1 NN and DL

2019年12月4日 23:52

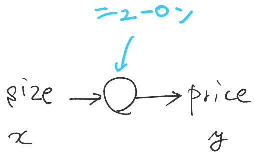
家賃予測

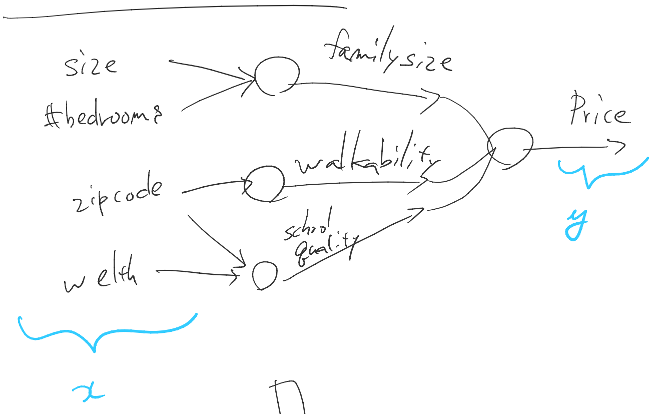


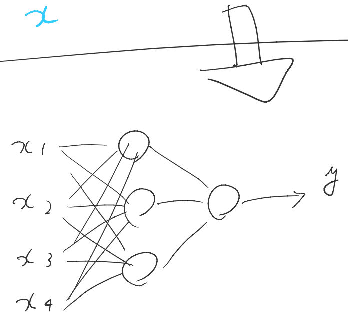
ReLU

Rectified Linear Unit

= max(0, val)







教師あり学習

CNN : 画像

RNN : 時系列、音声 など.

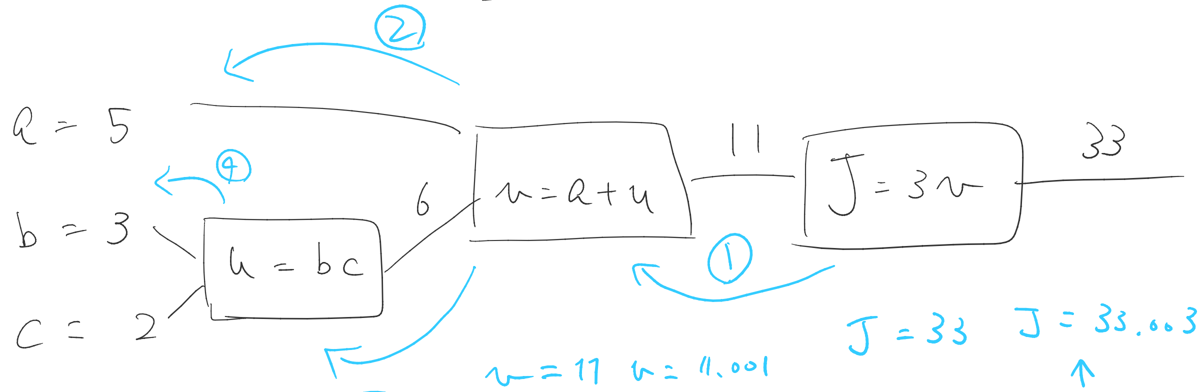
構造化データ : テーブル

非 〃 : 音声 画像, Text

week1 ここまで

Computation Graph

$$J(a,b,c) = 3(a+bc)$$
$$u = bc$$
$$v = a + u$$



$v = 11 \quad v = 11.001$
$a = 5 \quad a = 5.001$
$J = 33 \quad J = 33.003$

$$\frac{dJ}{da} = 3 = \frac{dJ}{dv}\frac{dv}{da}$$

$$\frac{dv}{da} = 1 \qquad 2$$

$J = 33 \quad J = 33.003$
$v = 11 \quad v = 11.001$

$$\frac{dJ}{dv} = 3 \qquad 1$$

$v = 11 \quad v = 11.001$
$u = 6 \quad u = 6.001$
$J = 33 \quad J = 33.003$

$$\frac{dJ}{du} = 3 = \frac{dJ}{dv}\frac{dv}{du} \qquad 3$$

$$\frac{dv}{du} = 1$$

$u = 6 \quad u = 6.002$
$b = 3 \quad b = 3.001$
$v = 11 \quad v = 11.002$
$J = 33 \quad J = 33.006$

4

$$\frac{du}{db} = 2$$

$$\frac{dv}{db} = \frac{dv}{du}\frac{du}{db} = 2$$

$$\frac{dJ}{db} = \frac{dJ}{dv}\frac{dv}{db} = 6$$

Logistic Regression

$$z = w^T x + b$$

$$z = w^T x + b$$

$$\hat{y} = \sigma(z) = a$$

$$\mathcal{L}(\ ,y) = -(y\log(\hat{y}) + (1-y)\log(1-\hat{y}))$$

x_1, w_1, x_2, w_2, b → $z = w_1 x_1 + w_2 x_2 + b$ → $a = \sigma(z)$ → $\mathcal{L}(a,y)$

$$\text{'}dz\text{'} = \frac{d\mathcal{L}}{dz} = a - y$$

②

$$= \frac{d\mathcal{L}}{da}\frac{da}{dz}$$

$$= \left(\frac{y}{a} + \frac{y}{1-a}\right)\cdot a(1-a)$$

$$\text{'}da\text{'} = \frac{d\mathcal{L}(a,y)}{da}$$

①

$$= -\frac{y}{a} + \frac{y}{1-a}$$

$$\text{'}dw_1\text{'} = \frac{\partial \mathcal{L}}{\partial w_1} = x_1 dz$$

$$\text{'}dw_2\text{'} = \frac{\partial \mathcal{L}}{\partial w_2} = x_2 dz$$

$$\text{'}db\text{'} = dz$$

③

Repeat

$$\begin{cases} w_1 := w_1 - \alpha \frac{\partial \mathcal{L}}{\partial w_1} \\ w_2 := w_2 - \alpha \frac{\partial \mathcal{L}}{\partial w_2} \\ b := b - \alpha \frac{\partial \mathcal{L}}{\partial b} \end{cases}$$

$$w^T = \overset{m\times 1}{\begin{bmatrix} \\ \\ \end{bmatrix}} \quad x = \overset{n\times m}{\begin{bmatrix} (1) & (2) & \cdots & (m) \\ | & | & | & | \end{bmatrix}} \quad b = \overset{1\times m}{\begin{bmatrix} b & b & \cdots & b \end{bmatrix}}$$

$$\boxed{w^T x} = w^T x^{(1)} \quad w^T x^{(2)} \dots w^T x^{(m)}$$

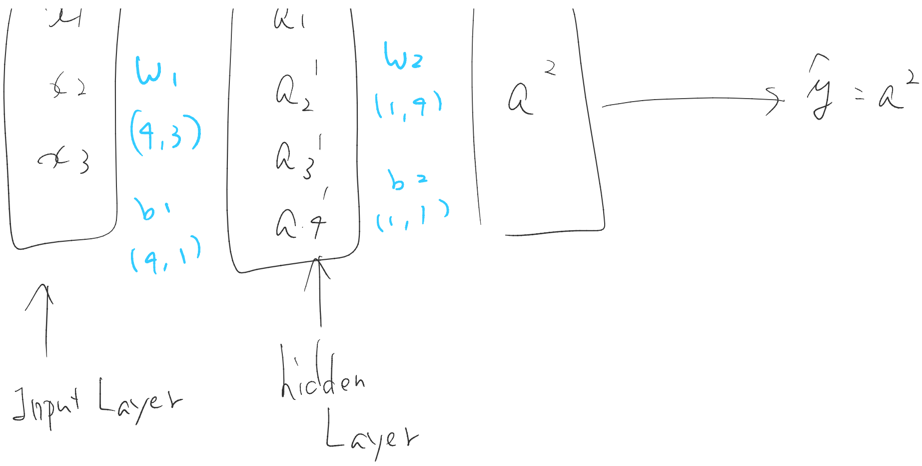
$$\boxed{w^T x + b} = w^T x^{(1)} + b \ \dots \ w^T x^{(m)} + b$$

Week2 ここまで

---

x_1 x_2 $\xrightarrow{w_1}$ a_1^1 a_2^1 $\xrightarrow{w_2}$ a^2 ⟶ $\hat{y} = a^2$

w_1 w_2 $w_1 z$ $w_2 z$



$W_1 x^1 + b_1 = a^1 \in \mathbb{R}^{4\times m}$

$x \in \mathbb{R}^{3\times m}$

$W_1 \in \mathbb{R}^{4\times 3}$

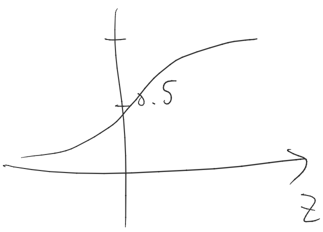
$b_1 \in \mathbb{R}^{4\times 1}$

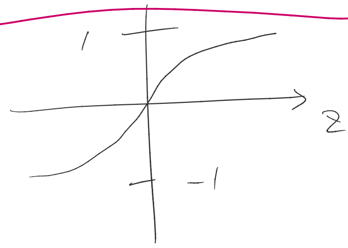
$W_2 a^1 + b_2 = a^2 \in \mathbb{R}^{1\times m}$

1,4 4,m 1,1

σ(z) あまり使われない

$\boxed{\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}}$





多くの場合, σ(z)より

tanh(z)がいい.

平均が0に近づくため

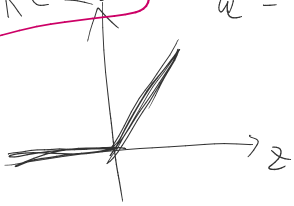
⇓

バイナリ分類の出力層は例外.

→ 0~1を出力したいため.

$\boxed{\text{ReLU}}$ よく使われる.

$a = \max(0, z)$



Leaky ReLU

$a = \max(0.01z, z)$

Activation function.

$g(z) = \sigma(z)$ ← これは何故そうするのか？

$g(z) = z$ 恒等関数 (1)=3) → ロジスティック回帰より性能が劣るだろう.

ただし、線形回帰の場合は出力層で使う

出力層が実数の

導関数

$$g(z) = \frac{1}{1+e^{-z}}$$

$$\frac{dg(z)}{dz} = \frac{1}{1+e^{-z}}\left(1 - \frac{1}{1+e^{-z}}\right)$$

$$= g(z)\left(1 - g(z)\right)$$

$z = 0 \rightarrow g(z) = 0.5$

$$\frac{dg(z)}{dz} = 0.5\,(1-0.5) = 0.25$$

$z = 10 \rightarrow g(z) \approx 1$

$$\frac{dg(z)}{dz} \approx 1\,(1-1) \approx 0$$

$z = -10 \rightarrow g(z) \approx 0$

$$\frac{dg(z)}{dz} \approx 0\,(1-0) \approx 0$$

$$g(z) = \tanh(z) = \frac{e^{-z} - e^{z}}{e^{z} + e^{z}}$$

$$g'(z) = 1 - (\tanh(z))^2$$

$z = 10$ $\tanh(z) \approx 1$
$g'(z) \approx 0$

$z = -10$ $\tanh(z) \approx -1$
$g'(z) \approx 0$

$z = 0$ $\tanh(z) = 0$
$g'(z) = 1$

ReLU

$$g(z) = \max(0, z)$$

$$g'(z) = \begin{cases} 0 & \text{if } z < 0 \\ 1 & \text{if } z \geq 0 \end{cases}$$

Leaky ReLU

$$g(z) = \max(0.001z, z)$$

$$g'(z) = \begin{cases} 0.001 & \text{if } z < 0 \\ 1 & \text{if } z \geq 0 \end{cases}$$

$g(z) = \begin{cases} 1 & \text{if } z \geq 0 \\ \cancel{\text{undefined}} & \cancel{\text{if } z < 0} \end{cases}$

$\begin{cases} 1 & \text{if } z \geq 0 \end{cases}$