$(x^{(i)}, y^{(i)})$ $\quad dw_1^{(i)}, dw_2^{(i)}, db^{(i)}$

$$\frac{\partial}{\partial w_1} J(w,b) = \frac{1}{m}\sum_{i=1}^{m} \underbrace{\frac{\partial}{\partial w_1} L(a^{(i)}, y^{(i)})}_{} = dw_1^{(i)}$$

$J=0,\quad dw_1=0,\quad dw_2=0\quad db=0$

For $i=1$ to m

$z^{(i)} = w^T x^{(i)} + b$

$a^{(i)} = \sigma(z^{(i)})$

$J += -[y^{(i)}\log a^{(i)} + (1-y^{(i)})\log(1-a^{(i)})]$

$dz^{(i)} = a^{(i)} - y^{(i)}$

$dw_1 += x_1^{(i)} dz^{(i)}$ $\quad n=2$

$dw_2 += x_2^{(i)} dz^{(i)}$

$db += dz^{(i)}$

$dw_3 \rightarrow \vdots\ dw_n$

$J /= m\ (J/m)\quad dw_1 /= m\ (dw_1/m)\quad dw_2 /= m\ (dw_2/m)\quad db /= m\ (db/m)$

$w_1 = w_1 - \alpha\, dw_1 \qquad w_2 := w_2 - \alpha\, dw_2 \qquad b := b - \alpha\, db$

四 向量化 / Vectorization)

$z = w^T x + b.\qquad w = \begin{bmatrix}\vdots\end{bmatrix} \in R^{n_x}\quad x = \begin{bmatrix}\vdots\end{bmatrix} \in R^{n_x}$

非向量化表示

$z = 0$

for $i=1$ in range(n_x)

$z += w[i] * x[i]$

$z += b\ (z = z + b)$

向量化

$z = \underbrace{np.dot(w, x)}_{w^T x} + b$

GPU } SIMD — single instruction multiple data
CPU }

$Z = [z^{(1)}, z^{(2)}, \ldots z^{(m)}]_{1\times m} = w^T X + [b, b, \ldots b]$

只要有可能，都应该尽量避免 for 循环. $\quad dZ = \begin{bmatrix} dz^{(1)} \\ \vdots \\ dz^{(m)}\end{bmatrix}$

$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \rightarrow u = \begin{bmatrix} e^{v_1} \\ e^{v_2} \\ \vdots \\ e^{v_n}\end{bmatrix}$

u = np.zeros((n,1))
for i in rang(n):
    u[i] = math.exp(v[i])

import numpy as np
u = np.exp(v)
np.log(v)
np.abs(v)

logistic 回归求导. $\quad dw = np.zeros((n_x, 1))\quad X = [x^{(1)}\ x^{(2)} \ldots x^{(m)}]\ R^{n_x \times m}$

$J=0,$ ~~dw1=0 dw2=0~~ $db=0$ $\quad$ for iteration in range(1000)

for $i=1$ to m: $\qquad Z = w^T X + b = np.dot(w.T, X) + b$

$z^{(i)} = w^T x^{(i)} + b$

$a^{(i)} = \sigma(z^{(i)})$

$J += -[y^{(i)}\log \hat{y}^{(i)} + (1-y^{(i)})\log(1-\hat{y}^{(i)})]$ $\quad A = \sigma(Z)$

$dz^{(i)} = a^{(i)}(1-a^{(i)})$ $\quad dZ = A - Y$

$dw += x^{(i)} dz^{(i)}$ $\quad dw = \frac{1}{m} X \cdot dZ^T$

$db += dz^{(i)}$ $\quad db = \frac{1}{m} np.sum(dZ)$

$J = J/m\quad dw = dw/m\quad db = db/m$ $\quad w = w - \alpha dw$

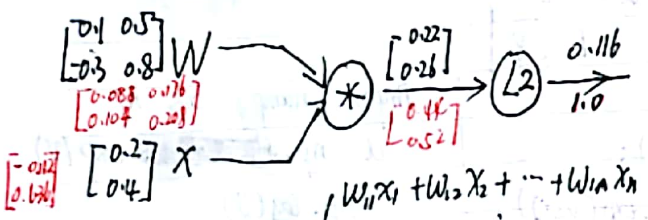
$b = b - \alpha db$

$\underset{[4096\times1]}{x} \rightarrow f(x) = \max(0, x) \xrightarrow{f} \underset{[4096\times1]}{}$

$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial f}\cdot \boxed{\frac{\partial f}{\partial x}}$ 雅可比矩阵，维数为4096×4096.

实际中当使用 mini-batch 运算时，(b=100)，则 Jacobian 矩阵为 [409600×409600] 维

例③
如 $f(x,W) = \|W\cdot x\|^2 = \sum_{i=1}^{n}(W\cdot x)_i^2 \quad x\in\mathbb{R}^n,\ W\in\mathbb{R}^{n\times n}$



$\nabla_x = 2W^T q$

$\nabla_W f = 2q\cdot x^T$

$$q = W\cdot x = \begin{pmatrix} W_{11}x_1 + W_{12}x_2 + \cdots + W_{1n}x_n \\ \vdots \\ W_{n1}x_1 + W_{n2}x_2 + \cdots + W_{nn}x_n \end{pmatrix}$$

$\frac{\partial f}{\partial q_i} = 2q_i \qquad \nabla_q f = 2q$

$f(q) = \|q\|^2 = q_1^2 + \cdots + q_n^2$

$\frac{\partial q_k}{\partial W_{ij}} = 1_{k=i}x_j$

$\frac{\partial f}{\partial W_{ij}} = \sum_k \frac{\partial f}{\partial q_k}\cdot\frac{\partial q_k}{\partial W_{ij}} = \sum_k (2q_k)(1_{k=i}x_j)$

$= 2q_i x_j$

$\frac{\partial q_k}{\partial x_i} = W_{k,i}$

$\frac{\partial f}{\partial x_i} = \sum_k \frac{\partial f}{\partial q_k}\cdot\frac{\partial q_k}{\partial x_i} = \sum_k 2q_k W_{k,i}$ 例3我到第二部分之前

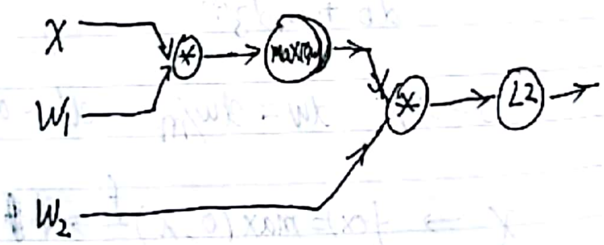
例④ 矩阵例子

$z_1 = XW_1$

$h_1 = ReLU(z_1)$

$\hat{y} = h_1 W_2$

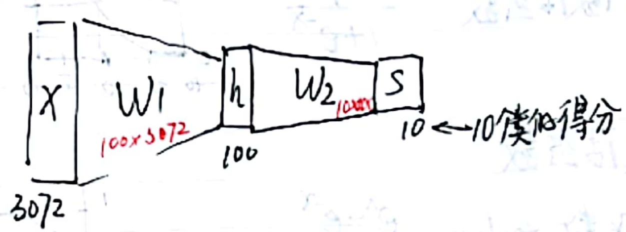
$L = \|\hat{y}\|_2^2$



五 神经网络 (Neural Networks)

线性得分函数 $f = Wx$

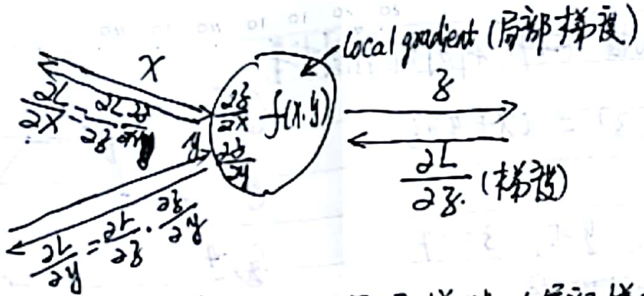
2-层神经网络 $f = W_2\max(0, W_1x)$



3层神经网络. $f = W_3\max(0, W_2\cdot\max(0, W_1x))$

训练一个两层神经网络需要~20行代码实现

```
import numpy as np
from numpy.random import randn
N, D_in, H, D_out = 64, 1000, 100, 10
x, y = randn(N, D_in), randn(N, D_out)
W1, W2 = randn(D_in, H), randn(H, D_out)
```

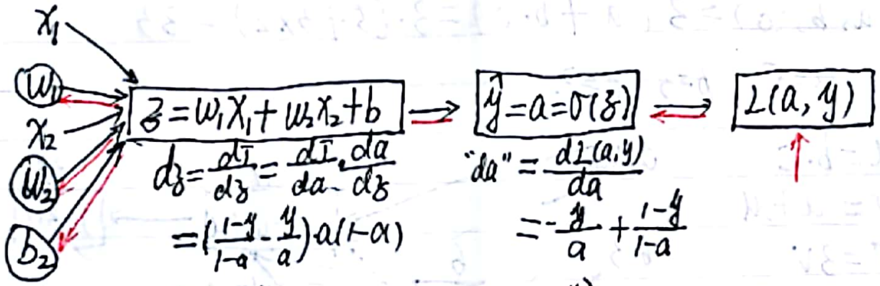


反馈梯度 = 下一层梯度 × 局部梯度

例3 ① logistic 回归

$$z = w^T x + b. \qquad \hat{y} = a = \sigma(z) = \frac{1}{1+e^{-z}}$$

$$J = L(a, y) = -\left(y\log a + (1-y)\log(1-a)\right)$$



$$dz = \frac{dJ}{dz} = \frac{dJ}{da}\cdot\frac{da}{dz} \qquad \text{“}da\text{”} = \frac{dL(a,y)}{da}$$

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

$$= a - y$$

$$dw_1 = \frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial z}\cdot\frac{\partial z}{\partial w_1} = x_1 \cdot dz = x_1(a-y) \qquad dw_2 = \frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial z}\cdot\frac{\partial z}{\partial w_2} = x_2 \cdot dz = x_2(a-y)$$

$$db = \frac{\partial L}{\partial b} = \frac{\partial L}{\partial z}\cdot\frac{\partial z}{\partial b} = dz = a - y$$

$$w_1 := w_1 - \alpha dw_1$$

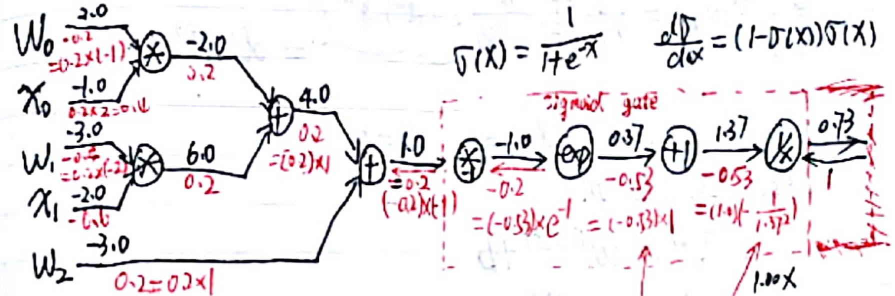
$$w_2 := w_2 - \alpha dw_2$$

$$b := b - \alpha db$$

例② 数值化的 Sigmoid 函数

$$f(w, x) = \frac{1}{1+e^{-(w_0x_0 + w_1x_1 + w_2)}}$$

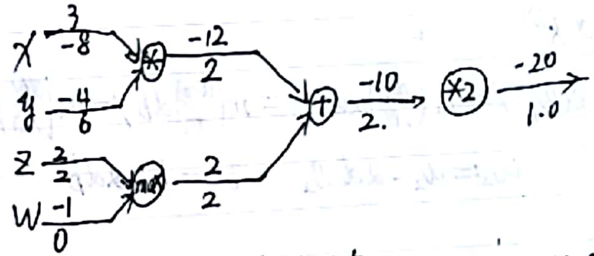
$$\sigma(x) = \frac{1}{1+e^{-x}} \qquad \frac{d\sigma}{dx} = (1-\sigma(x))\sigma(x)$$



$$f(x) = e^x \rightarrow \frac{df}{dx} = e^x \qquad f(x) = \frac{1}{x} \rightarrow \frac{df}{dx} = -1/x^2$$

$$f_a(x) = ax \rightarrow \frac{df}{dx} = a \qquad f_c(x) = c + x \rightarrow \frac{df}{dx} = 1$$

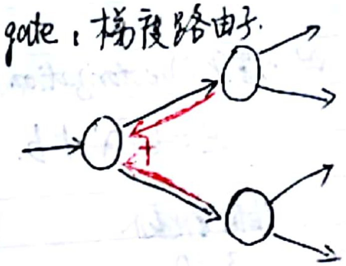
$$J = 2 \times (x \cdot y + \max\{w, z\})$$



add gate : 梯度分布子.

mul gate : 梯度变换子

max gate : 梯度路由子.



反向传播时, 梯度为各个分支梯度相加。

三. 多个样本的 logistic 回归

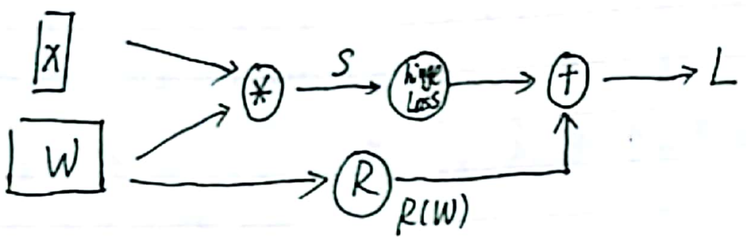
$$J(w,b) = \frac{1}{m}\sum_{i=1}^{m} L(a^{(i)}, y^{(i)})$$

$$a^{(i)} = \hat{y}^{(i)} = \sigma(z^{(i)}) = \sigma(w^T x^{(i)} + b)$$

第四讲 反向传播和神经网络

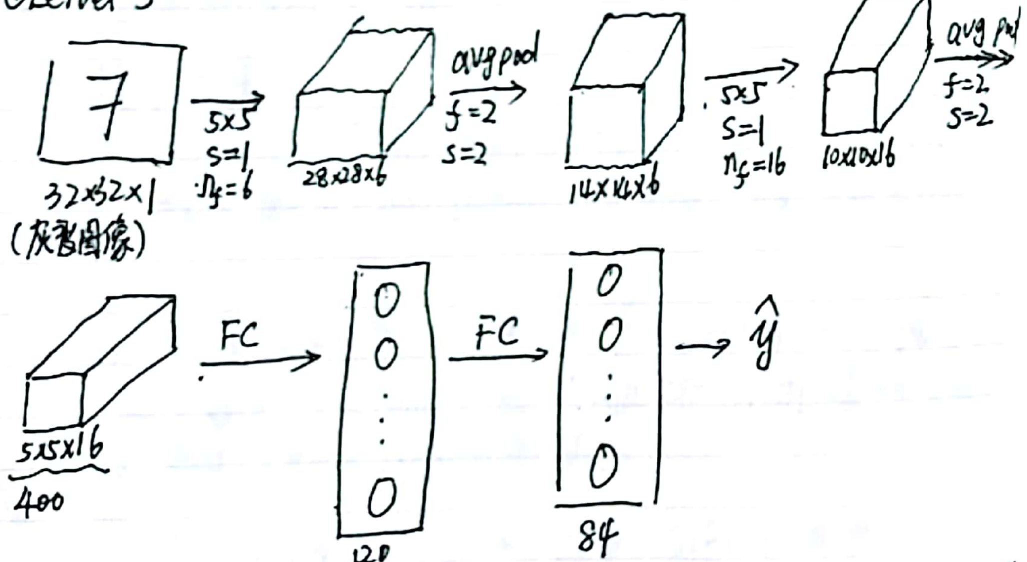
一 计算图. ~~(利用计算图求导)~~

$f = Wx$ $\quad L_i = \sum_{j \neq y_i} \max(0, S_j - S_{y_i} + 1)$.



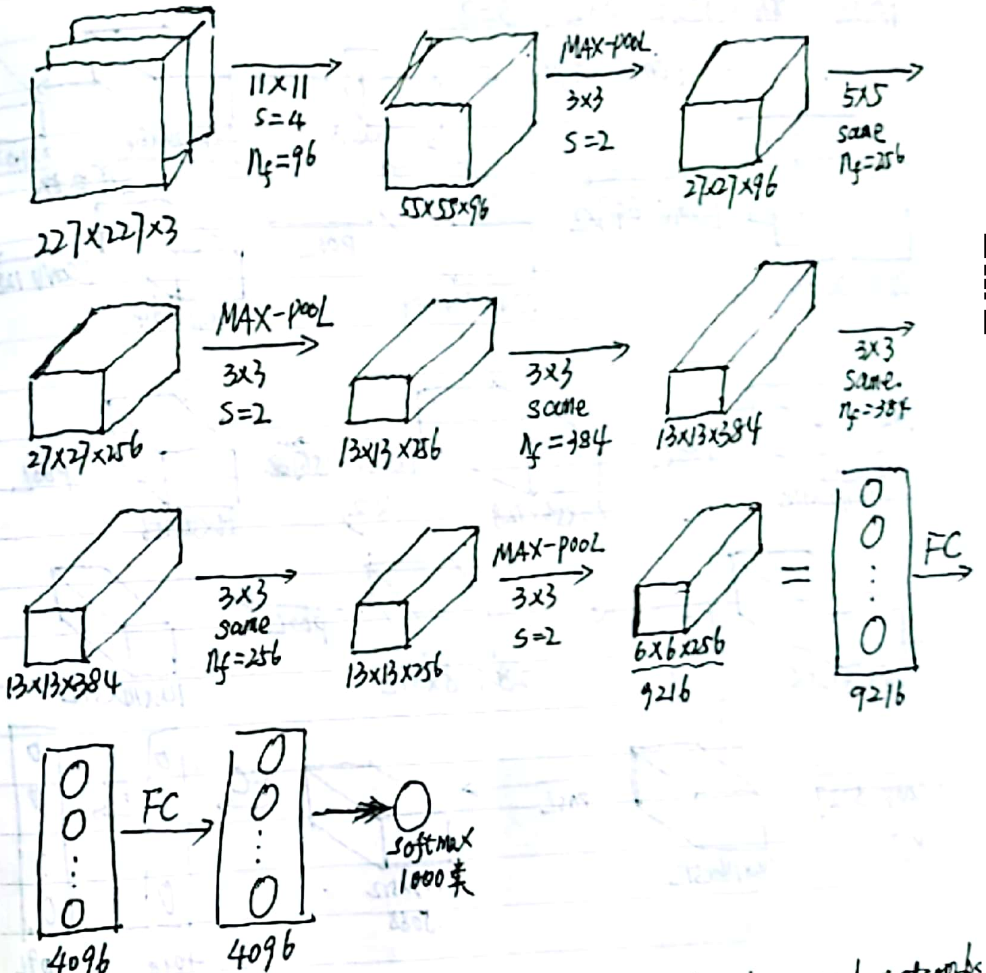
Convolutional network (AlexNet) 网络结构. LeNet-5 and VGG11

①LeNet-5



[LeCun 1998. Gradient-based learning applied to document recognition]

② AlexNet

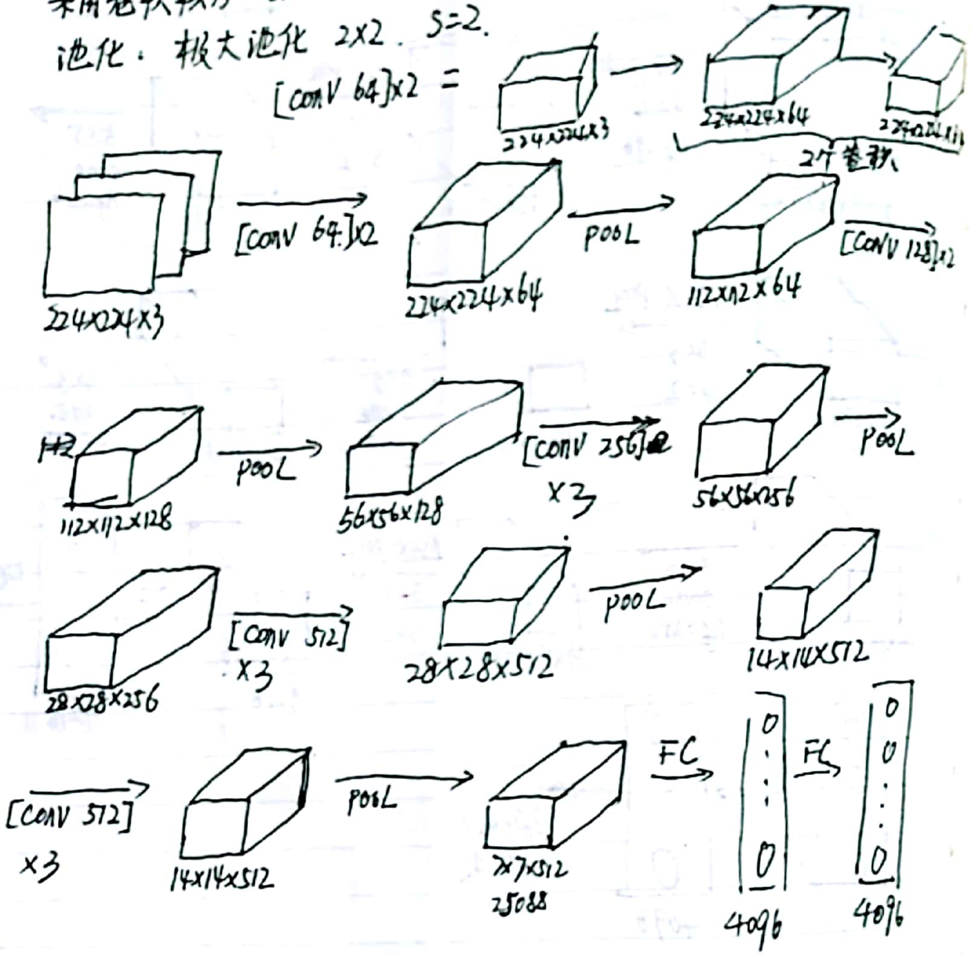


[ImageNet Classification with deep convolutional neural networks
2012 krizhevsky]

VGG-16

采用卷积核为 conv=3x3. $s=1$. same.;

池化：极大池化 2x2. $s=2$.

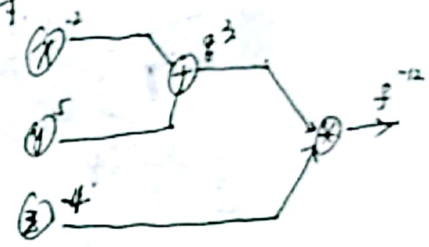


→ softmax
1000

[Simonyan & Zisserman 2015 Very deep convolution networks for large-scale image recognition

E. 反向传播 ~~(二)~~ 利用计算图求导

$f(x,y,z) = (x+y)\cdot z$



eg. $x=-2$, $y=5$, $z=-4$

~~$q=3$, $f=-12$~~

$q = x+y = 3$ $\qquad \frac{\partial q}{\partial x}=1$. $\qquad \frac{\partial q}{\partial y}=1$

$f = qz$ $\qquad \frac{\partial f}{\partial z}=q$ $\qquad \frac{\partial f}{\partial q}=z$.

目标 $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, $\frac{\partial f}{\partial z}$.

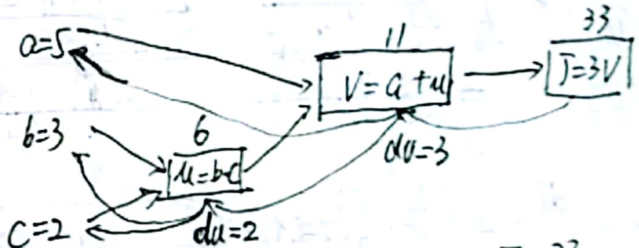
$J(a,b,c) = 3(a+b\cdot c) = 3\cdot(5+3\times2) = 33$

$a=5$, $b=3$ $c=2$

$u=b\cdot c$

$V=a+u$

$J=3V$



$J=3V$ $\qquad V=11$. $\qquad J=33$

$V=11.001$ $\qquad J=33.003$

$a=5.001$ $\qquad V=11.001$ $\qquad J=33.003$

$\frac{dJ}{da}$, $\frac{dJ}{db}$ $\frac{dJ}{dc}$

$\frac{dJ}{dV}=3$ $\qquad \frac{dJ}{da}=3$

$=\frac{dJ}{dV}\cdot\frac{dV}{da}=3$

$u=6 \to 6.001$ $V=11.001$ $J=33.003$

$b=3\to3.001$ $u=6.002$. $V=11.002$ $J=33.006$

$\frac{dJ}{du}=3$ $\qquad \frac{dJ}{db}=6$

$\frac{dJ}{da}=\frac{dJ}{du}\cdot\frac{du}{da}=3\times3=9$.