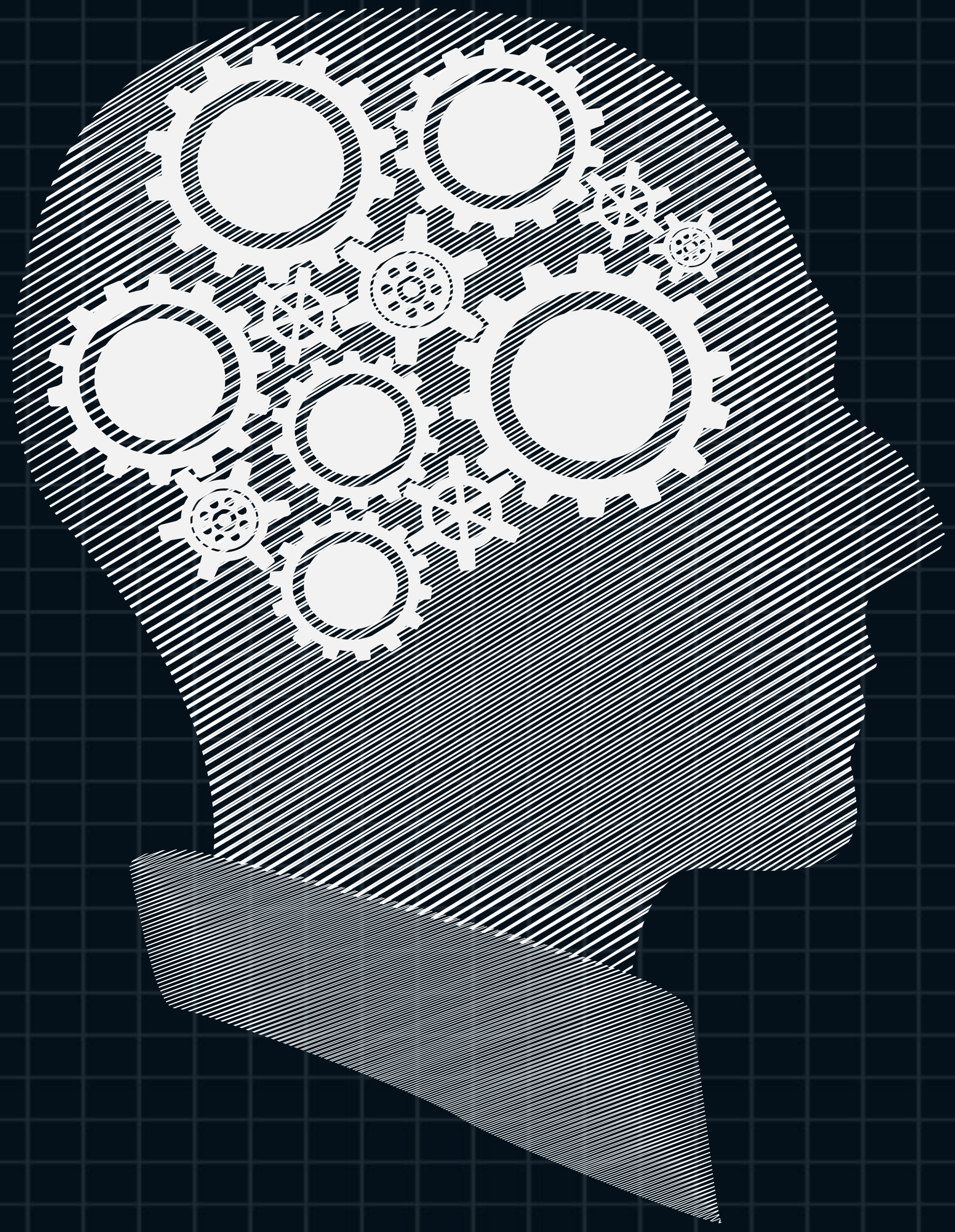


Backpropagation and Jacobian Matrices

Lecture.7
Linear/Logistic Regression(2)

Lecture.7 Linear/Logistic Regression(2) - Linear Regression

Model and Tensors

$$J = \frac{1}{N}\sum_{i=1}^{N} J_0^{(i)}$$

$$\vec{J_0} = \begin{pmatrix} \left(y^{(1)} - \hat{y}^{(1)}\right)^2 \\ \left(y^{(2)} - \hat{y}^{(2)}\right)^2 \\ \vdots \\ \left(y^{(N)} - \hat{y}^{(N)}\right)^2 \end{pmatrix} = \begin{pmatrix} J_0^{(1)} \\ J_0^{(2)} \\ \vdots \\ J_0^{(N)} \end{pmatrix}$$

J

$$J = \mathscr{L}\left(\vec{y}, \vec{\hat{y}}\right)$$

$\vec{\hat{y}}$

$$\vec{\hat{y}} = X^T \cdot \vec{w} + b$$

X

$$\vec{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}, \ b$$

$$\vec{\hat{y}} = X^T \cdot \vec{w} + b = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_{l_I}^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \dots & x_{l_I}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(N)} & x_2^{(N)} & \dots & x_{l_I}^{(N)} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_{l_I} \end{pmatrix} + b = \begin{pmatrix} \hat{y}^{(1)} \\ \hat{y}^{(2)} \\ \vdots \\ \hat{y}^{(N)} \end{pmatrix}$$

$$X^T = \begin{pmatrix} \longleftarrow & \left(\vec{x}^{(1)}\right)^T & \longrightarrow \\ \longleftarrow & \left(\vec{x}^{(2)}\right)^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \left(\vec{x}^{(N)}\right)^T & \longrightarrow \end{pmatrix} = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_{l_I}^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \dots & x_{l_I}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(N)} & x_2^{(N)} & \dots & x_{l_I}^{(N)} \end{pmatrix}$$

Jacobians

$$J = \frac{1}{N}\sum_{i=1}^{N} J_0^{(i)}$$

$$\frac{\partial J}{\partial \overrightarrow{J}_0} = \begin{pmatrix} \frac{\partial J}{\partial J_0^{(1)}} & \frac{\partial J}{\partial J_0^{(2)}} & \cdots & \frac{\partial J}{\partial J_0^{(N)}} \end{pmatrix} = \begin{pmatrix} \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \end{pmatrix}$$

$$\overrightarrow{J_0} = \begin{pmatrix} \left(y^{(1)} - \hat{y}^{(1)}\right)^2 \\ \left(y^{(2)} - \hat{y}^{(2)}\right)^2 \\ \vdots \\ \left(y^{(N)} - \hat{y}^{(N)}\right)^2 \end{pmatrix} = \begin{pmatrix} J_0^{(1)} \\ J_0^{(2)} \\ \vdots \\ J_0^{(N)} \end{pmatrix}$$

$$\frac{\partial \overrightarrow{J}_0}{\partial \overrightarrow{\hat{y}}} = \begin{pmatrix} -2\left(y^{(1)} - \hat{y}^{(1)}\right) & 0 & \dots & 0 \\ 0 & -2\left(y^{(2)} - \hat{y}^{(2)}\right) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -2\left(y^{(N)} - \hat{y}^{(N)}\right) \end{pmatrix}$$

$$\overrightarrow{\hat{y}} = X \cdot \overrightarrow{w} + b = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_{l_I}^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \dots & x_{l_I}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(N)} & x_2^{(N)} & \dots & x_{l_I}^{(N)} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_{l_I} \end{pmatrix} + b = \begin{pmatrix} \hat{y}^{(1)} \\ \hat{y}^{(2)} \\ \vdots \\ \hat{y}^{(N)} \end{pmatrix}$$

$$\frac{\partial \overrightarrow{\hat{y}}}{\partial \overrightarrow{w}} = \begin{pmatrix} \longleftarrow & \left(\overrightarrow{x}^{(1)}\right)^T & \longrightarrow \\ \longleftarrow & \left(\overrightarrow{x}^{(2)}\right)^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \left(\overrightarrow{x}^{(N)}\right)^T & \longrightarrow \end{pmatrix} = X^T$$

$$X^T = \begin{pmatrix} \longleftarrow & \left(\overrightarrow{x}^{(1)}\right)^T & \longrightarrow \\ \longleftarrow & \left(\overrightarrow{x}^{(2)}\right)^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \left(\overrightarrow{x}^{(N)}\right)^T & \longrightarrow \end{pmatrix} = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_{l_I}^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \dots & x_{l_I}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(N)} & x_2^{(N)} & \dots & x_{l_I}^{(N)} \end{pmatrix}$$

$$\frac{\partial \overrightarrow{\hat{y}}}{\partial b} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

Lecture.7 Linear/Logistic Regression(2) - Linear Regression

Backpropagation

$$\frac{\partial J}{\partial \vec{J}_0} = \begin{pmatrix} \frac{\partial J}{\partial J_0^{(1)}} & \frac{\partial J}{\partial J_0^{(2)}} & \cdots & \frac{\partial J}{\partial J_0^{(N)}} \end{pmatrix} = \begin{pmatrix} \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \end{pmatrix}$$

$$\frac{\partial \vec{J}_0}{\partial \vec{\hat{y}}} = \begin{pmatrix} -2\left(y^{(1)} - \hat{y}^{(1)}\right) & 0 & \dots & 0 \\ 0 & -2\left(y^{(2)} - \hat{y}^{(2)}\right) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -2\left(y^{(N)} - \hat{y}^{(N)}\right) \end{pmatrix}$$

$$\frac{\partial J}{\partial \vec{\hat{y}}} = \frac{\partial J}{\partial \vec{J}_0}\frac{\partial \vec{J}_0}{\partial \vec{\hat{y}}} = -\frac{2}{N}\left((y^{(1)} - \hat{y}^{(1)}) \quad (y^{(2)} - \hat{y}^{(2)}) \quad \dots \quad (y^{(N)} - \hat{y}^{(N)})\right)$$

$$\frac{\partial \vec{\hat{y}}}{\partial \vec{w}} = \begin{pmatrix} \longleftarrow & \left(\vec{x}^{(1)}\right)^T & \longrightarrow \\ \longleftarrow & \left(\vec{x}^{(2)}\right)^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \left(\vec{x}^{(N)}\right)^T & \longrightarrow \end{pmatrix} = X^T$$

$$\frac{\partial J}{\partial \vec{w}} = \frac{\partial J}{\partial \vec{\hat{y}}}\frac{\partial \vec{\hat{y}}}{\partial \vec{w}} = -\frac{2}{N}\left((y^{(1)} - \hat{y}^{(1)}) \quad (y^{(2)} - \hat{y}^{(2)}) \quad \dots \quad (y^{(N)} - \hat{y}^{(N)})\right)\begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_{l_x}^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \dots & x_{l_x}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(N)} & x_2^{(N)} & \dots & x_{l_x}^{(N)} \end{pmatrix}$$

$$= \frac{-2}{N}\left(\sum_{i=1}^N x_1^{(i)}(y^{(i)} - \hat{y}^{(i)}) \quad \sum_{i=1}^N x_2^{(i)}(y^{(i)} - \hat{y}^{(i)}) \quad \dots \quad \sum_{i=1}^N x_{l_x}^{(i)}(y^{(i)} - \hat{y}^{(i)})\right)$$

$$\frac{\partial \vec{\hat{y}}}{\partial b} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$\frac{\partial J}{\partial b} = \frac{\partial J}{\partial \vec{\hat{y}}}\frac{\partial \vec{\hat{y}}}{\partial b} = -\frac{2}{N}\left((y^{(1)} - \hat{y}^{(1)}) \quad (y^{(2)} - \hat{y}^{(2)}) \quad \dots \quad (y^{(N)} - \hat{y}^{(N)})\right)\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$= \frac{-2}{N}\sum_{i=1}^N \left(y^{(i)} - \hat{y}^{(i)}\right)$$

Parameter Update

$$\frac{\partial J}{\partial \vec{w}} = \frac{-2}{N}\left(\sum_{i=1}^{N} x_1^{(i)}(y^{(i)} - \hat{y}^{(i)}) \quad \sum_{i=1}^{N} x_2^{(i)}(y^{(i)} - \hat{y}^{(i)}) \quad \dots \quad \sum_{i=1}^{N} x_{l_x}^{(i)}(y^{(i)} - \hat{y}^{(i)}) \right)$$

$$\frac{\partial J}{\partial b} = \frac{-2}{N}\sum_{i=1}^{N}\left(y^{(i)} - \hat{y}^{(i)}\right)$$

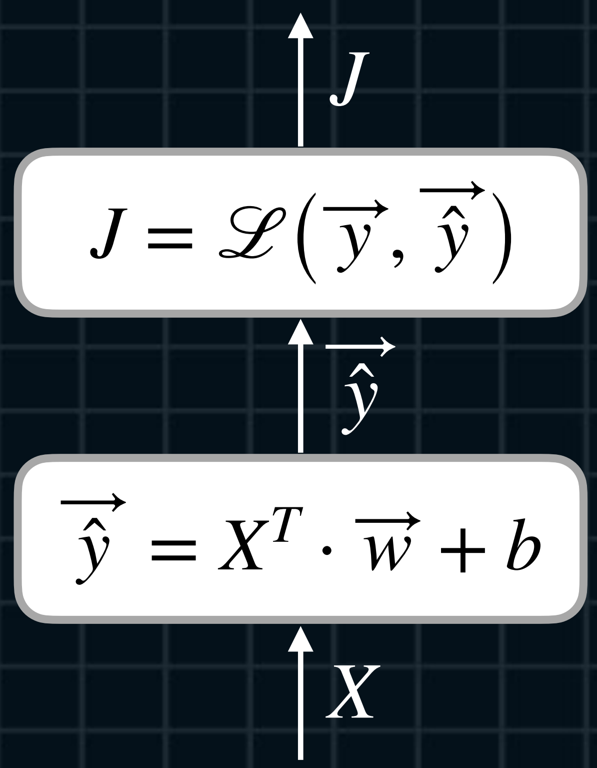
$$\vec{w} := \vec{w} - \alpha\left(\frac{\partial J}{\partial \vec{w}}\right)^T$$

$$b := b - \alpha\frac{\partial J}{\partial b}$$

$$w_j := w_j + \frac{2\alpha}{N}\sum_{i=1}^{N} x_j^{(i)}(y^{(i)} - \hat{y}^{(i)})$$

$$b := b + \frac{2\alpha}{N}\sum_{i=1}^{N}\left(y^{(i)} - \hat{y}^{(i)}\right)$$

Backpropagation with Matrices



$$\frac{\partial J}{\partial \vec{J}_0} = \begin{pmatrix} \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \end{pmatrix}$$

$$\frac{\partial \vec{J}_0}{\partial \hat{\vec{y}}} = \begin{pmatrix} -2\left(y^{(1)} - \hat{y}^{(1)}\right) & 0 & \cdots & 0 \\ 0 & -2\left(y^{(2)} - \hat{y}^{(2)}\right) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -2\left(y^{(N)} - \hat{y}^{(N)}\right) \end{pmatrix}$$

$$\frac{\partial \hat{\vec{y}}}{\partial \vec{w}} = \begin{pmatrix} \longleftarrow & \left(\vec{x}^{(1)}\right)^T & \longrightarrow \\ \longleftarrow & \left(\vec{x}^{(2)}\right)^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \left(\vec{x}^{(N)}\right)^T & \longrightarrow \end{pmatrix} = X^T$$

$$\frac{\partial \hat{\vec{y}}}{\partial b} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$\underset{(1, l_I)}{\frac{\partial J}{\partial \vec{w}}} = \underset{(1,N)}{\frac{\partial J}{\partial \vec{J}_0}} \underset{(N,N)}{\frac{\partial \vec{J}_0}{\partial \hat{\vec{y}}}} \underset{(N, l_I)}{\frac{\partial \hat{\vec{y}}}{\partial \vec{w}}}$$

$$\underset{(1,1)}{\frac{\partial J}{\partial b}} = \underset{(1,N)}{\frac{\partial J}{\partial \vec{J}_0}} \underset{(N,N)}{\frac{\partial \vec{J}_0}{\partial \hat{\vec{y}}}} \underset{(N,1)}{\frac{\partial \hat{\vec{y}}}{\partial b}}$$

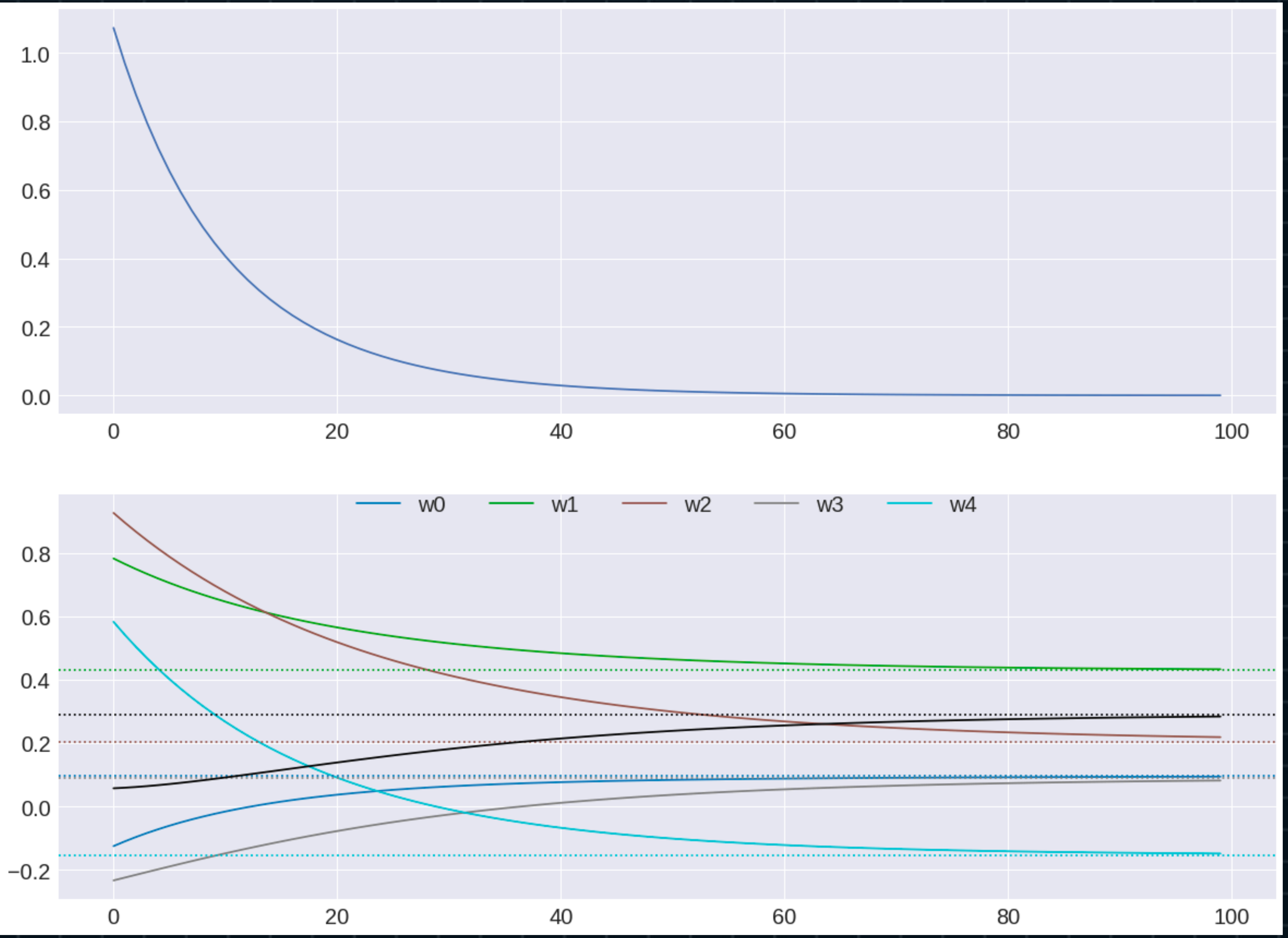
Implementation

```
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.cm as cm
np.random.seed(0)
plt.style.use('seaborn')

# set params
N, n_feature = 32, 5
lr = 0.03
t_W = np.random.uniform(-1, 1, n_feature).reshape(-1, 1)
t_b = np.random.uniform(-1, 1, 1)
W = np.random.uniform(-1, 1, n_feature).reshape(-1, 1)
b = np.random.uniform(-1, 1, 1).reshape(1, 1)
epochs = 100


# generate dataset
X = np.random.randn(N, n_feature)
Y = X @ t_W + t_b

J_list = list()
W_list, b_list = list(), list()
```

```
for epoch in range(epochs):
  W_list.append(W)
  b_list.append(b)

  # loss calculation
  Pred = X @ W + b
  J0 = (Y - Pred)**2
  J = np.mean(J0)
  J_list.append(J)

  # jacobians
  dJ_dJ0 = 1/N*np.ones((1, N))
  dJ0_dPred = np.diag(-2*(Y - Pred).flatten())
  dPred_dW = X
  dPred_db = np.ones((N, 1))

  # backpropagation
  dJ_dPred = dJ_dJ0 @ dJ0_dPred
  dJ_dW = dJ_dPred @ dPred_dW
  dJ_db = dJ_dPred @ dPred_db

  # parameter update
  W = W - lr*dJ_dW.T
  b = b - lr*dJ_db

W_list = np.hstack(W_list)
b_list = np.concatenate(b_list)
```

Implementation

```python
# visualize results
cmap = cm.get_cmap('tab10', n_feature)
fig, axes = plt.subplots(2, 1, figsize=(20, 15))
axes[0].plot(J_list)

for w_idx, w_list in enumerate(W_list):
  axes[1].plot(w_list, color=cmap(w_idx), label='w' + str(w_idx))
for w_idx, t_w in enumerate(t_W):
  axes[1].axhline(y=t_w, linestyle=':', color=cmap(w_idx))
axes[1].plot(b_list, color='black')
axes[1].axhline(y=t_b, linestyle=':', color='black')
axes[1].legend(fontsize=20, loc='lower center',
               bbox_to_anchor=(0.5, 0.9), ncol=n_feature)
axes[0].tick_params(labelsize=20)
axes[1].tick_params(labelsize=20)
```



Model and Tensors

$$
\begin{array}{c}
\uparrow J \\
\boxed{J = \mathscr{L}\left(\overrightarrow{y}, \overrightarrow{\hat{y}}\right)} \\
\uparrow \overrightarrow{\hat{y}} \\
\boxed{\overrightarrow{\hat{y}} = \sigma\left(\vec{z}\right)} \\
\uparrow \vec{z} \\
\boxed{\vec{z} = X^T \cdot \overrightarrow{w} + b} \\
\uparrow X
\end{array}
$$

$$J = \frac{1}{N}\sum_{i=1}^{N} J_0^{(i)}$$

$$\overrightarrow{J_0} = \begin{pmatrix} y^{(1)}log\left(\hat{y}^{(1)}\right) + (1 - y^{(1)})log\left(1 - \hat{y}^{(1)}\right) \\ y^{(2)}log\left(\hat{y}^{(2)}\right) + (1 - y^{(2)})log\left(1 - \hat{y}^{(2)}\right) \\ \vdots \\ y^{(N)}log\left(\hat{y}^{(N)}\right) + (1 - y^{(N)})log\left(1 - \hat{y}^{(N)}\right) \end{pmatrix} = \begin{pmatrix} J_0^{(1)} \\ J_0^{(2)} \\ \vdots \\ J_0^{(N)} \end{pmatrix}$$

$$\overrightarrow{a} = \sigma\left(\vec{z}\right) = \begin{pmatrix} \sigma\left(z^{(1)}\right) \\ \sigma\left(z^{(2)}\right) \\ \vdots \\ \sigma\left(z^{(N)}\right) \end{pmatrix} = \begin{pmatrix} a^{(1)} \\ a^{(2)} \\ \vdots \\ a^{(N)} \end{pmatrix}$$

$$\vec{z} = X^T \cdot \overrightarrow{w} + b = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_{l_I}^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \dots & x_{l_I}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(N)} & x_2^{(N)} & \dots & x_{l_I}^{(N)} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_{l_I} \end{pmatrix} + b = \begin{pmatrix} z^{(1)} \\ z^{(2)} \\ \vdots \\ z^{(N)} \end{pmatrix}$$

$$X^T = \begin{pmatrix} \longleftarrow & \left(\overrightarrow{x}^{(1)}\right)^T & \longrightarrow \\ \longleftarrow & \left(\overrightarrow{x}^{(2)}\right)^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \left(\overrightarrow{x}^{(N)}\right)^T & \longrightarrow \end{pmatrix} = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_{l_I}^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \dots & x_{l_I}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(N)} & x_2^{(N)} & \dots & x_{l_I}^{(N)} \end{pmatrix}$$

- Logistic Regression

Jacobians

$$J = \frac{1}{N}\sum_{i=1}^{N} J_0^{(i)}$$

$$\vec{J_0} = \begin{pmatrix} y^{(1)}log\left(\hat{y}^{(1)}\right) + (1-y^{(1)})log\left(1-\hat{y}^{(1)}\right) \\ y^{(2)}log\left(\hat{y}^{(2)}\right) + (1-y^{(2)})log\left(1-\hat{y}^{(2)}\right) \\ \vdots \\ y^{(N)}log\left(\hat{y}^{(N)}\right) + (1-y^{(N)})log\left(1-\hat{y}^{(N)}\right) \end{pmatrix} = \begin{pmatrix} J_0^{(1)} \\ J_0^{(2)} \\ \vdots \\ J_0^{(N)} \end{pmatrix}$$

$$\vec{a} = \sigma(\vec{z}) = \begin{pmatrix} \sigma\left(z^{(1)}\right) \\ \sigma\left(z^{(2)}\right) \\ \vdots \\ \sigma\left(z^{(N)}\right) \end{pmatrix} = \begin{pmatrix} a^{(1)} \\ a^{(2)} \\ \vdots \\ a^{(N)} \end{pmatrix}$$

$$\vec{z} = X^T \cdot \vec{w} + b = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_{l_I}^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \dots & x_{l_I}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(N)} & x_2^{(N)} & \dots & x_{l_I}^{(N)} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_{l_I} \end{pmatrix} + b = \begin{pmatrix} z^{(1)} \\ z^{(2)} \\ \vdots \\ z^{(N)} \end{pmatrix}$$

$$X^T = \begin{pmatrix} \longleftarrow & \left(\vec{x}^{(1)}\right)^T & \longrightarrow \\ \longleftarrow & \left(\vec{x}^{(2)}\right)^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \left(\vec{x}^{(N)}\right)^T & \longrightarrow \end{pmatrix} = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_{l_I}^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \dots & x_{l_I}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(N)} & x_2^{(N)} & \dots & x_{l_I}^{(N)} \end{pmatrix}$$

$$\frac{\partial J}{\partial \vec{J}_0} = \begin{pmatrix} \frac{\partial J}{\partial J_0^{(1)}} & \frac{\partial J}{\partial J_0^{(2)}} & \cdots & \frac{\partial J}{\partial J_0^{(N)}} \end{pmatrix} = \begin{pmatrix} \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \end{pmatrix}$$

$$\frac{\partial \vec{J}_0}{\partial \vec{\hat{y}}} = \begin{pmatrix} \frac{\hat{y}^{(1)}-y^{(1)}}{\hat{y}^{(1)}(1-\hat{y}^{(1)})} & 0 & \dots & 0 \\ 0 & \frac{\hat{y}^{(2)}-y^{(2)}}{\hat{y}^{(2)}(1-\hat{y}^{(2)})} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\hat{y}^{(N)}-y^{(N)}}{\hat{y}^{(N)}(1-\hat{y}^{(N)})} \end{pmatrix}$$

$$\frac{\partial \vec{\hat{y}}}{\partial \vec{z}} = \begin{pmatrix} \hat{y}^{(1)}\left(1-\hat{y}^{(1)}\right) & 0 & \dots & 0 \\ 0 & \hat{y}^{(2)}\left(1-\hat{y}^{(2)}\right) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{y}^{(N)}\left(1-\hat{y}^{(N)}\right) \end{pmatrix}$$

$$\frac{\partial \vec{z}}{\partial \vec{w}} = \begin{pmatrix} \longleftarrow & \left(\vec{x}^{(1)}\right)^T & \longrightarrow \\ \longleftarrow & \left(\vec{x}^{(2)}\right)^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \left(\vec{x}^{(N)}\right)^T & \longrightarrow \end{pmatrix} = X^T \quad \frac{\partial \vec{z}}{\partial b} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

- Logistic Regression

Backpropagation

$$\frac{\partial J}{\partial \vec{J}_0} = \begin{pmatrix} \frac{\partial J}{\partial J_0^{(1)}} & \frac{\partial J}{\partial J_0^{(2)}} & \cdots & \frac{\partial J}{\partial J_0^{(N)}} \end{pmatrix} = \begin{pmatrix} \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \end{pmatrix}$$

$$\frac{\partial \vec{J}_0}{\partial \vec{\hat{y}}} = \begin{pmatrix} \frac{\hat{y}^{(1)} - y^{(1)}}{\hat{y}^{(1)}(1-\hat{y}^{(1)})} & 0 & \cdots & 0 \\ 0 & \frac{\hat{y}^{(2)} - y^{(2)}}{\hat{y}^{(2)}(1-\hat{y}^{(2)})} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\hat{y}^{(N)} - y^{(N)}}{\hat{y}^{(N)}(1-\hat{y}^{(N)})} \end{pmatrix}$$

$$\frac{\partial \vec{\hat{y}}}{\partial \vec{z}} = \begin{pmatrix} \hat{y}^{(1)}\left(1-\hat{y}^{(1)}\right) & 0 & \cdots & 0 \\ 0 & \hat{y}^{(2)}\left(1-\hat{y}^{(2)}\right) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{y}^{(N)}\left(1-\hat{y}^{(N)}\right) \end{pmatrix}$$

$$\frac{\partial \vec{z}}{\partial \vec{w}} = \begin{pmatrix} \longleftarrow & \left(\vec{x}^{(1)}\right)^T & \longrightarrow \\ \longleftarrow & \left(\vec{x}^{(2)}\right)^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \left(\vec{x}^{(N)}\right)^T & \longrightarrow \end{pmatrix} = X^T \qquad \frac{\partial \vec{z}}{\partial b} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$\frac{\partial J}{\partial \vec{\hat{y}}} = \frac{1}{N}\begin{pmatrix} \frac{\hat{y}^{(1)} - y^{(1)}}{\hat{y}^{(1)}(1-\hat{y}^{(1)})} & \frac{\hat{y}^{(2)} - y^{(2)}}{\hat{y}^{(2)}(1-\hat{y}^{(2)})} & \cdots & \frac{\hat{y}^{(N)} - y^{(N)}}{\hat{y}^{(N)}(1-\hat{y}^{(N)})} \end{pmatrix}$$

$$\begin{aligned} \frac{\partial J}{\partial \vec{z}} &= \frac{\partial J}{\partial \vec{\hat{y}}}\frac{\partial \vec{\hat{y}}}{\partial \vec{z}} = \frac{1}{N}\begin{pmatrix} \frac{\hat{y}^{(1)} - y^{(1)}}{\hat{y}^{(1)}(1-\hat{y}^{(1)})} & \frac{\hat{y}^{(2)} - y^{(2)}}{\hat{y}^{(2)}(1-\hat{y}^{(2)})} & \cdots & \frac{\hat{y}^{(N)} - y^{(N)}}{\hat{y}^{(N)}(1-\hat{y}^{(N)})} \end{pmatrix} \begin{pmatrix} \hat{y}^{(1)}\left(1-\hat{y}^{(1)}\right) & 0 & \cdots & 0 \\ 0 & \hat{y}^{(2)}\left(1-\hat{y}^{(2)}\right) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{y}^{(N)}\left(1-\hat{y}^{(N)}\right) \end{pmatrix} \\ &= \frac{1}{N}\begin{pmatrix} \hat{y}^{(1)} - y^{(1)} & \hat{y}^{(2)} - y^{(2)} & \cdots & \hat{y}^{(N)} - y^{(N)} \end{pmatrix} \end{aligned}$$

$$\begin{aligned} \frac{\partial J}{\partial \vec{w}} &= \frac{\partial J}{\partial \vec{z}}\frac{\partial \vec{z}}{\partial \vec{w}} = \frac{1}{N}\begin{pmatrix} \hat{y}^{(1)} - y^{(1)} & \hat{y}^{(2)} - y^{(2)} & \cdots & \hat{y}^{(N)} - y^{(N)} \end{pmatrix} \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_{l_x}^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_{l_x}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(N)} & x_2^{(N)} & \cdots & x_{l_x}^{(N)} \end{pmatrix} \\ &= -\frac{1}{N}\begin{pmatrix} \sum_{i=1}^N x_1^{(i)}(y^{(i)} - \hat{y}^{(i)}) & \sum_{i=1}^N x_2^{(i)}(y^{(i)} - \hat{y}^{(i)}) & \cdots & \sum_{i=1}^N x_{l_x}^{(i)}(y^{(i)} - \hat{y}^{(i)}) \end{pmatrix} \end{aligned}$$

$$\begin{aligned} \frac{\partial J}{\partial b} &= \frac{\partial J}{\partial \vec{z}}\frac{\partial \vec{z}}{\partial b} = -\frac{1}{N}\begin{pmatrix} (y^{(1)} - \hat{y}^{(1)}) & (y^{(2)} - \hat{y}^{(2)}) & \cdots & (y^{(N)} - \hat{y}^{(N)}) \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \\ &= -\frac{1}{N}\sum_{i=1}^N \left(y^{(i)} - \hat{y}^{(i)}\right) \end{aligned}$$

Parameter Update

$$\frac{\partial J}{\partial \vec{w}} = -\frac{1}{N}\left(\sum_{i=1}^{N} x_1^{(i)}(y^{(i)} - \hat{y}^{(i)}) \quad \sum_{i=1}^{N} x_2^{(i)}(y^{(i)} - \hat{y}^{(i)}) \quad \dots \quad \sum_{i=1}^{N} x_{l_x}^{(i)}(y^{(i)} - \hat{y}^{(i)}) \right)$$

$$\frac{\partial J}{\partial b} = -\frac{1}{N}\sum_{i=1}^{N}\left(y^{(i)} - \hat{y}^{(i)}\right)$$

$$\vec{w} := \vec{w} - \alpha\left(\frac{\partial J}{\partial \vec{w}}\right)^{T}$$

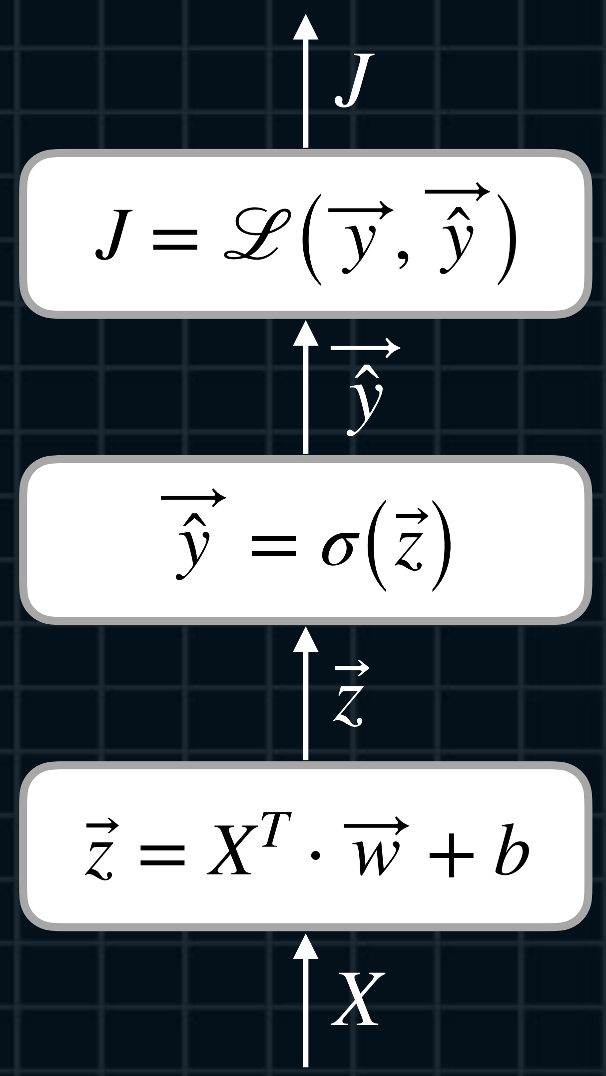
$$b := b - \alpha\frac{\partial J}{\partial b}$$

$$w_j := w_j + \frac{\alpha}{N}\sum_{i=1}^{N} x_j^{(i)}(y^{(i)} - \hat{y}^{(i)})$$

$$b := b + \frac{\alpha}{N}\sum_{i=1}^{N}\left(y^{(i)} - \hat{y}^{(i)}\right)$$

Lecture.7 Linear/Logistic Regression(2) - Logistic Regression

Backpropagation with Matrices



$$\frac{\partial J}{\partial \vec{J}_0} = \begin{pmatrix} \frac{\partial J}{\partial J_0^{(1)}} & \frac{\partial J}{\partial J_0^{(2)}} & \cdots & \frac{\partial J}{\partial J_0^{(N)}} \end{pmatrix} = \begin{pmatrix} \frac{1}{N} & \frac{1}{N} & \dots & \frac{1}{N} \end{pmatrix}$$

$$\frac{\partial \vec{J}_0}{\partial \vec{\hat{y}}} = \begin{pmatrix} \frac{\hat{y}^{(1)} - y^{(1)}}{\hat{y}^{(1)}(1-\hat{y}^{(1)})} & 0 & \dots & 0 \\ 0 & \frac{\hat{y}^{(2)} - y^{(2)}}{\hat{y}^{(2)}(1-\hat{y}^{(2)})} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\hat{y}^{(N)} - y^{(N)}}{\hat{y}^{(N)}(1-\hat{y}^{(N)})} \end{pmatrix}$$

$$\frac{\partial \vec{\hat{y}}}{\partial \vec{z}} = \begin{pmatrix} \hat{y}^{(1)}(1-\hat{y}^{(1)}) & 0 & \dots & 0 \\ 0 & \hat{y}^{(2)}(1-\hat{y}^{(2)}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{y}^{(N)}(1-\hat{y}^{(N)}) \end{pmatrix}$$

$$\frac{\partial \vec{z}}{\partial \vec{w}} = \begin{pmatrix} \longleftarrow & (\vec{x}^{(1)})^T & \longrightarrow \\ \longleftarrow & (\vec{x}^{(2)})^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & (\vec{x}^{(N)})^T & \longrightarrow \end{pmatrix} = X^T \qquad \frac{\partial \vec{z}}{\partial b} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$\underset{(1,l_I)}{\frac{\partial J}{\partial \vec{w}}} = \underset{(1,N)}{\frac{\partial J}{\partial \vec{J}_0}} \underset{(N,N)}{\frac{\partial \vec{J}_0}{\partial \vec{\hat{y}}}} \underset{(N,N)}{\frac{\partial \vec{\hat{y}}}{\partial \vec{z}}} \underset{(N,l_I)}{\frac{\partial \vec{z}}{\partial \vec{w}}}$$

$$\underset{(1,1)}{\frac{\partial J}{\partial b}} = \underset{(1,N)}{\frac{\partial J}{\partial \vec{J}_0}} \underset{(N,N)}{\frac{\partial \vec{J}_0}{\partial \vec{\hat{y}}}} \underset{(N,N)}{\frac{\partial \vec{\hat{y}}}{\partial \vec{z}}} \underset{(N,1)}{\frac{\partial \vec{z}}{\partial b}}$$

- Logistic Regression

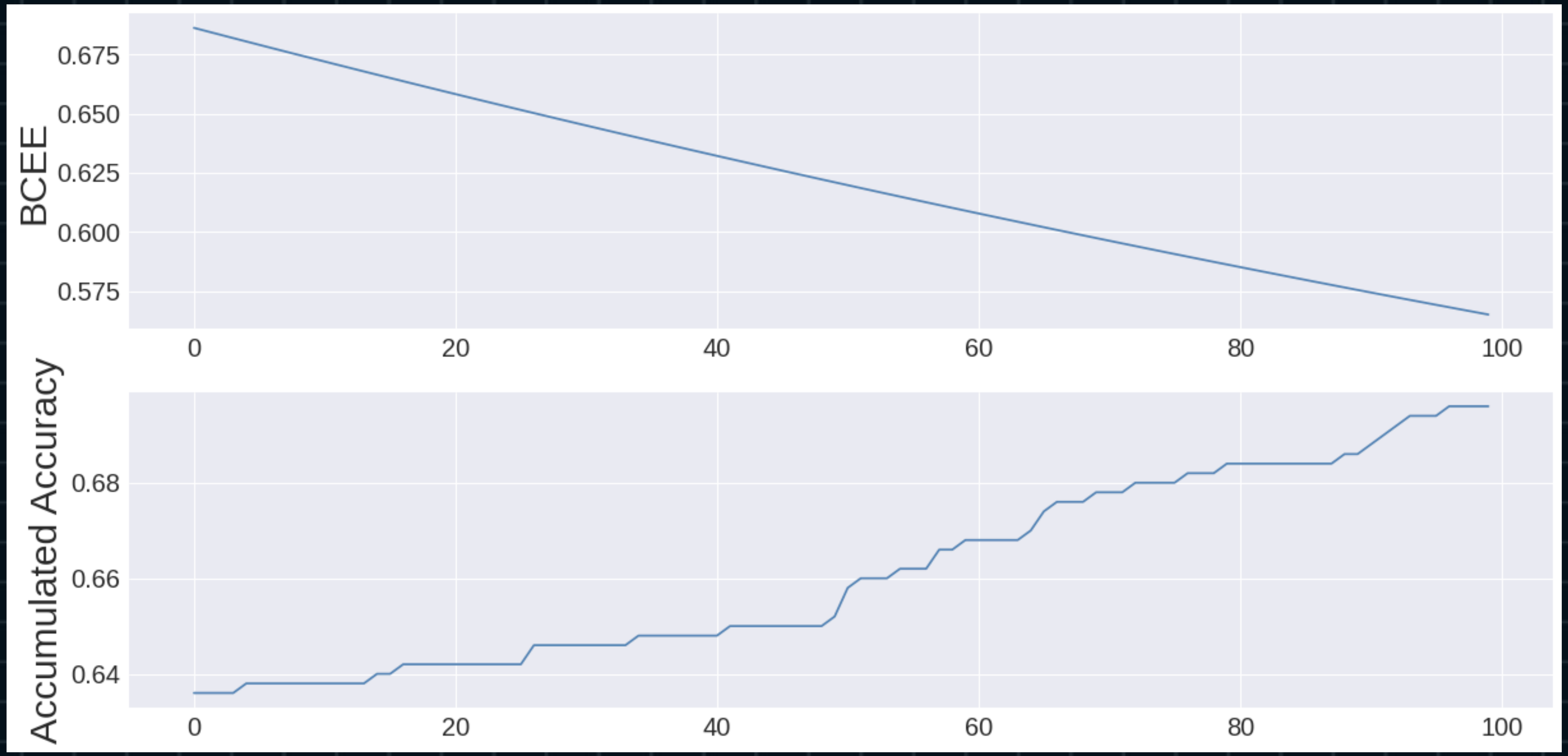
Implementation

```python
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.cm as cm
np.random.seed(1)
plt.style.use('seaborn')

# set params
N, n_feature = 500, 3
lr = 0.01
t_W = np.random.uniform(-1, 1, n_feature).reshape(-1, 1)
t_b = np.random.uniform(-1, 1, 1)
W = np.random.uniform(-1, 1, n_feature).reshape(-1, 1)
b = np.random.uniform(-1, 1, 1).reshape(1, 1)
epochs = 100


# generate dataset
X = np.random.normal(0, 1, (N, n_feature))
Y = X @ t_W + t_b
Y = (Y > 0).astype(np.int)

J_track = list()
acc_track = list()
```

```python
for epoch in range(epochs):
  # forward Ppropagation
  Z = X @ W + b
  Pred = 1/(1 + np.exp(-Z))
  J0 = -(Y*np.log(Pred) + (1-Y)*np.log(1-Pred))
  J = np.mean(J0)
  J_track.append(J)

  # calculate accuracy
  Pred_ = (Pred > 0.5).astype(np.int)
  n_correct = (Pred_ == Y).astype(np.int)
  acc = np.sum(n_correct)/N
  acc_track.append(acc)

  # jacobians
  dJ_dJ0 = 1/N*np.ones((1, N))
  dJ0_dPred = np.diag(((Pred - Y)/(Pred*(1-Pred))).flatten())
  dPred_dZ = np.diag((Pred*(1-Pred)).flatten())
  dZ_dW = X
  dZ_db = np.ones((N, 1))

  # backpropagation
  dJ_dPred = dJ_dJ0 @ dJ0_dPred
  dJ_dZ = dJ_dPred @ dPred_dZ
  dJ_dW = dJ_dZ @ dZ_dW
  dJ_db = dJ_dZ @ dZ_db

  # parameter update
  W = W - lr*dJ_dW.T
  b = b - lr*dJ_db
```

Implementation

```python
# visualize loss
fig, axes = plt.subplots(2, 1, figsize=(20, 10))
axes[0].plot(J_track)
axes[0].set_ylabel('BCEE', fontsize=30)
axes[0].tick_params(labelsize=20)

axes[1].plot(acc_track)
axes[1].tick_params(labelsize=20)
axes[1].set_ylabel('Accumulated Accuracy', fontsize=30)
```



Backpropagation and Jacobian Matrices

Lecture.7
Linear/Logistic Regression(2)