Logistic Regression

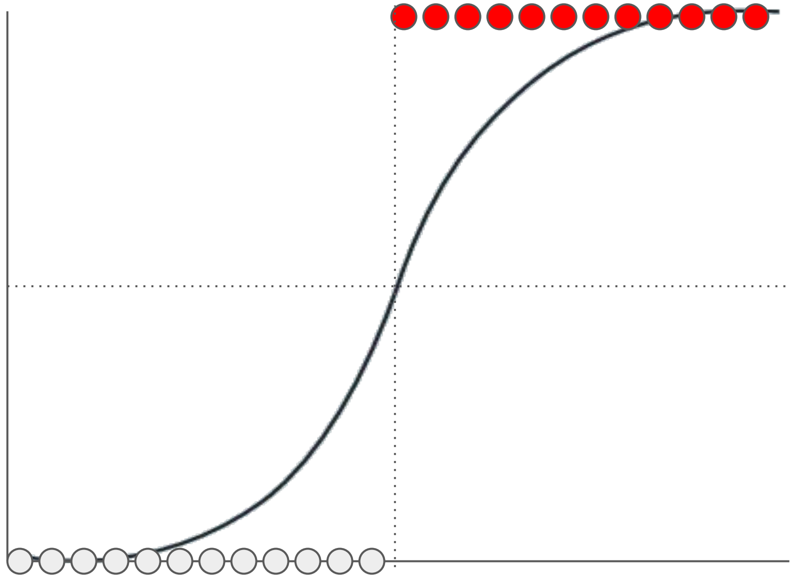
MACHINE LEARNING

Pakarat Musikawan

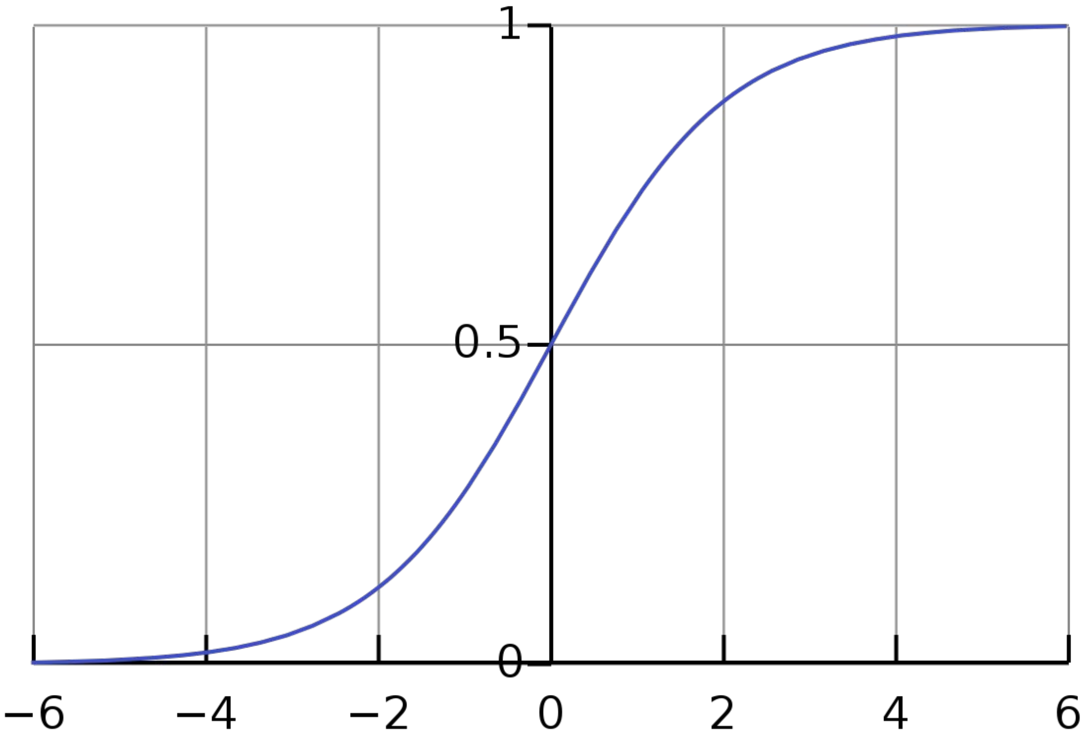
Logistic Regression vs Linear Regression

	Linear Regression	Logistic Regression
Definition	To predict a continuous dependent output variable (Y) based on independent input variables (X)	To predict a categorical dependent output variable (Y) based on independent input variables (X)
Variable type	Continuous dependent variable	Categorical dependent variable
Estimation method	Least square estimation	Maximum likelihood estimation
Equation	$\hat{y} = w_0 + \sum_{i=1}^{n} x_i w_i$	$\hat{y} = \frac{1}{1 + e^{-(w_0 + \sum_{i=1}^{n} x_i w_i)}}$
Best fit line	Straight line	Curve
Output	Predicted continuous value	Predicted probability [0, 1]

Logistic Regression



Logistic Regression



$$f(x) = \frac{1}{1 + \exp(-x)}$$

$$1 - f(x) = f(-x)$$

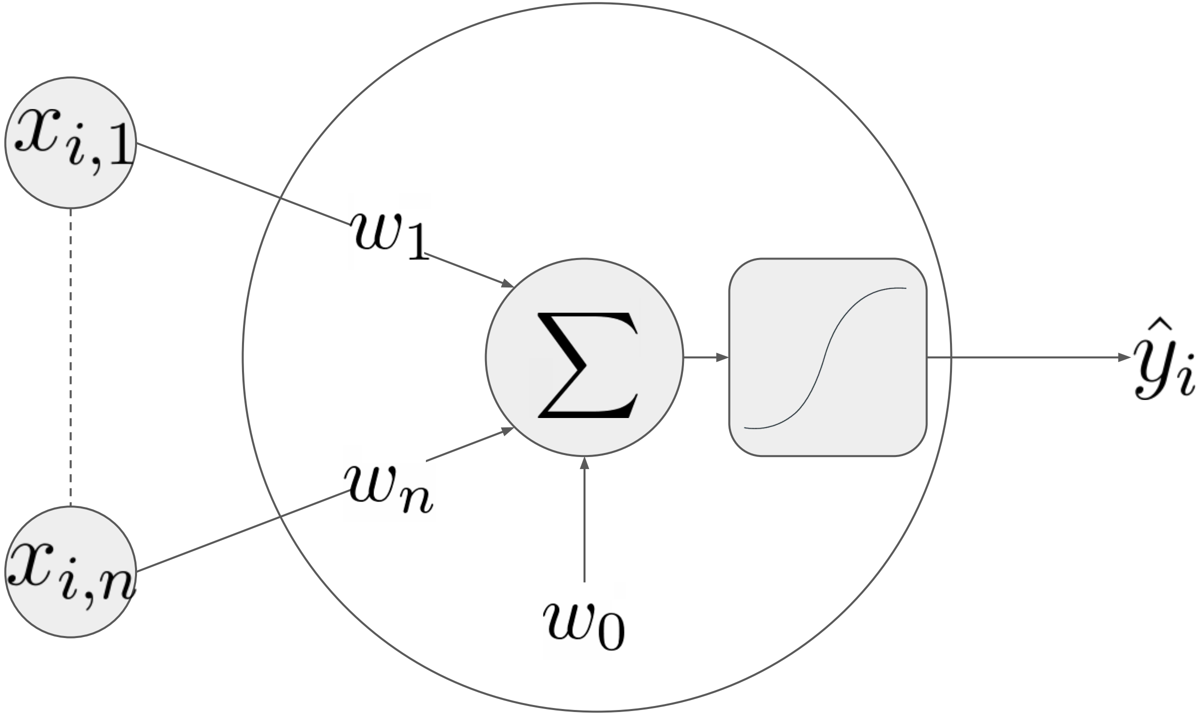
$$f(x) + f(-x) = 1$$

$$f(\inf) \to 1$$

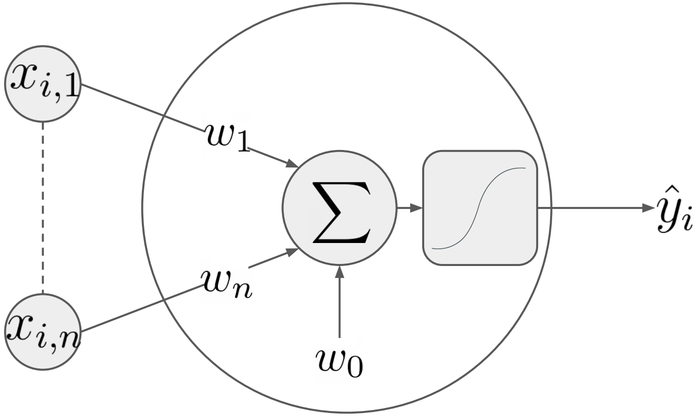
$$f(-\inf) \to 0$$

$$f(x) \in (0, 1), x \in (-\inf, \inf)$$

Logistic Regression



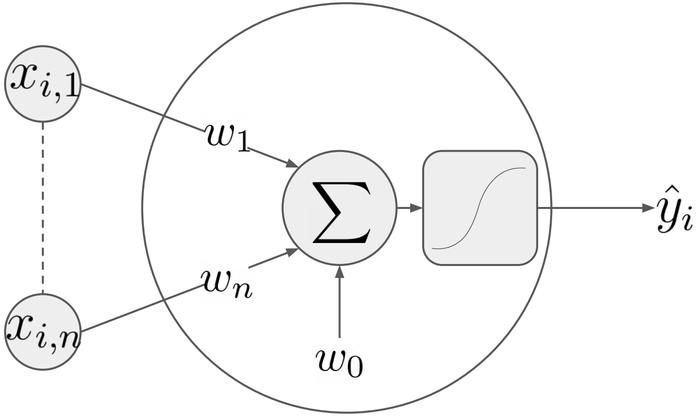
Logistic Regression



$\hat{y}_i$
0.1
0.6
0.8
0.4
0.7

$$d(\hat{y}_i) = \begin{cases} 1, & \text{if } \hat{y}_i \geq 0.5 \\ 0, & \text{if } \hat{y}_i < 0.5 \end{cases}$$

Logistic Regression



$1-\hat{y}_i$	$\hat{y}_i$
0.9	0.1
0.4	0.6
0.2	0.8
0.6	0.4
0.3	0.7

$$d(\hat{\boldsymbol{y}}_i) = \arg\max_k \hat{\boldsymbol{y}}_i$$

$$\hat{\boldsymbol{y}}_i = [\hat{y}_i, 1 - \hat{y}_i]$$

$$d([0.9, 0.1]) = 0$$

$$d([0.4, 0.6]) = 1$$

$$d([0.8, 0.2]) = 0$$

$$d([0.6, 0.4]) = 0$$

$$d([0.3, 0.7]) = 1$$

Parameters learning

$$\hat{y}_i = \frac{1}{1 + e^{-(w_0 + \sum_{j=1}^{n} x_{i,j} w_j)}}$$

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} -y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i)$$

Parameters learning

$$\hat{y}_i = \frac{1}{1 + e^{-(w_0 + \sum_{j=1}^{n} x_{i,j} w_j)}}$$

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} -y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i)$$

$$\boxed{w_j \leftarrow w_j - \eta \frac{\partial \mathcal{L}}{\partial w_j}}$$

Parameters learning

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} -y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i)$$

$$\hat{y}_i = \frac{1}{1 + e^{-z_i}}$$

$$z_i = w_0 + \sum_{j=1}^{n} x_{i,j} w_j$$

Parameters learning

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} -y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i)$$

$$\hat{y}_i = \frac{1}{1 + e^{-z_i}}$$

$$z_i = w_0 + \sum_{j=1}^{n} x_{i,j} w_j$$

$$\boxed{\frac{\partial \mathcal{L}}{\partial w_j} = \frac{\partial \mathcal{L}}{\partial \hat{y}_i} \frac{\partial \hat{y}_i}{\partial z_i} \frac{\partial z_i}{\partial w_j}}$$

Parameters learning $\dfrac{\partial \mathcal{L}}{\partial \hat{y}_i}$

$$\frac{\partial \mathcal{L}}{\partial \hat{y}_i} = \frac{\partial}{\partial \hat{y}_i}\frac{1}{N}\sum_{i=1}^{N} -y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)$$

$$\boxed{\begin{aligned}
\frac{\partial}{\partial x} f(x)g(x) &= f(x)\frac{\partial}{\partial x}g(x) + g(x)\frac{\partial}{\partial x}f(x) \\
&= \left[\frac{1}{N}\frac{\partial}{\partial \hat{y}_i}\sum_{i=1}^{N} -y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)\right] + \left[\sum_{i=1}^{N} -y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)\frac{\partial}{\partial \hat{y}_i}\frac{1}{N}\right] \\
&= \frac{1}{N}\frac{\partial}{\partial \hat{y}_i}\sum_{i=1}^{N} -y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)
\end{aligned}}$$

Parameters learning $\frac{\partial \mathcal{L}}{\partial \hat{y}_i}$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \hat{y}_i} &= \frac{\partial}{\partial \hat{y}_i}\frac{1}{N}\sum_{i=1}^{N} -y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i) \\ &= \frac{1}{N}\frac{\partial}{\partial \hat{y}_i}\sum_{i=1}^{N} -y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)\end{aligned}$$

$$\boxed{\begin{aligned}\frac{\partial}{\partial x}(f(x)+g(x)) &= \frac{\partial}{\partial x}f(x) + \frac{\partial}{\partial x}g(x) \\ \frac{\partial}{\partial x}\sum_i f_i(x) &= \sum_i \frac{\partial}{\partial x}f_i(x) \\ &= \frac{1}{N}\frac{\partial}{\partial \hat{y}_i}\sum_{i=1}^{N} -y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i) \\ &= \frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial \hat{y}_i} - (y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)) \\ &= \frac{1}{N}\sum_{i=1}^{N} -\frac{\partial}{\partial \hat{y}_i}(y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i))\end{aligned}}$$

Parameters learning $\dfrac{\partial \mathcal{L}}{\partial \hat{y}_i}$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \hat{y}_i} &= \frac{\partial}{\partial \hat{y}_i}\frac{1}{N}\sum_{i=1}^{N} -y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i) \\
&= \frac{1}{N}\frac{\partial}{\partial \hat{y}_i}\sum_{i=1}^{N} -y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i) \\
&= \frac{1}{N}\sum_{i=1}^{N} -\frac{\partial}{\partial \hat{y}_i}\left(y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)\right)
\end{aligned}
$$

$$
\boxed{
\begin{aligned}
\frac{\partial}{\partial x}\left(f(x)+g(x)\right) &= \frac{\partial}{\partial x}f(x) + \frac{\partial}{\partial x}g(x) \\
&= \frac{1}{N}\sum_{i=1}^{N} -\frac{\partial}{\partial \hat{y}_i}\left(y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)\right) \\
&= \frac{1}{N}\sum_{i=1}^{N} -\frac{\partial}{\partial \hat{y}_i}y_i \log \hat{y}_i + \frac{\partial}{\partial \hat{y}_i}(1-y_i)\log(1-\hat{y}_i)
\end{aligned}
}
$$

Parameters learning $\frac{\partial \mathcal{L}}{\partial \hat{y}_i}$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \hat{y}_i} &= \frac{\partial}{\partial \hat{y}_i} \frac{1}{N} \sum_{i=1}^{N} -y_i \log \hat{y}_i + (1-y_i) \log(1-\hat{y}_i) \\
&= \frac{1}{N} \frac{\partial}{\partial \hat{y}_i} \sum_{i=1}^{N} -y_i \log \hat{y}_i + (1-y_i) \log(1-\hat{y}_i) \\
&= \frac{1}{N} \sum_{i=1}^{N} -\frac{\partial}{\partial \hat{y}_i} \left(y_i \log \hat{y}_i + (1-y_i) \log(1-\hat{y}_i)\right) \\
&= \frac{1}{N} \sum_{i=1}^{N} -\frac{\partial}{\partial \hat{y}_i} y_i \log \hat{y}_i + \frac{\partial}{\partial \hat{y}_i} (1-y_i) \log(1-\hat{y}_i)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial x} f(x) g(x) &= f(x) \frac{\partial}{\partial x} g(x) + g(x) \frac{\partial}{\partial x} f(x) \\
\frac{\partial}{\partial \hat{y}_i} y_i \log \hat{y}_i &= y_i \frac{\partial}{\partial \hat{y}_i} \log \hat{y}_i + \log \hat{y}_i \frac{\partial}{\partial \hat{y}_i} y_i \\
&= y_i \frac{\partial}{\partial \hat{y}_i} \log \hat{y}_i
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial x} \log x &= \frac{1}{x} \\
\frac{\partial}{\partial \hat{y}_i} \log \hat{y}_i &= \frac{1}{\hat{y}_i}
\end{aligned}
$$

Parameters learning $\frac{\partial \mathcal{L}}{\partial \hat{y}_i}$

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \hat{y}_i} &= \frac{\partial}{\partial \hat{y}_i} \frac{1}{N} \sum_{i=1}^{N} -y_i \log \hat{y}_i + (1-y_i) \log (1-\hat{y}_i) \\
&= \frac{1}{N} \frac{\partial}{\partial \hat{y}_i} \sum_{i=1}^{N} -y_i \log \hat{y}_i + (1-y_i) \log (1-\hat{y}_i) \\
&= \frac{1}{N} \sum_{i=1}^{N} -\frac{\partial}{\partial \hat{y}_i} \left(y_i \log \hat{y}_i + (1-y_i) \log (1-\hat{y}_i)\right) \\
&= \frac{1}{N} \sum_{i=1}^{N} -\frac{\partial}{\partial \hat{y}_i} y_i \log \hat{y}_i + \frac{\partial}{\partial \hat{y}_i} (1-y_i) \log (1-\hat{y}_i) \\
&= \frac{1}{N} \sum_{i=1}^{N} -\frac{y_i}{\hat{y}_i} + \frac{\partial}{\partial \hat{y}_i} (1-y_i) \log (1-\hat{y}_i)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial x} f(x) g(x) &= f(x) \frac{\partial}{\partial x} g(x) + g(x) \frac{\partial}{\partial x} f(x) \\
&= y_i \frac{\partial}{\partial \hat{y}_i} \log \hat{y}_i \\
\frac{\partial}{\partial \hat{y}_i} (1-y_i) \log (1-\hat{y}_i) &= (1-y_i) \frac{\partial}{\partial \hat{y}_i} \log (1-\hat{y}_i) + \log (1-\hat{y}_i) \frac{\partial}{\partial \hat{y}_i} (1-y_i) \\
&= (1-y_i) \frac{\partial}{\partial \hat{y}_i} \log (1-\hat{y}_i)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial x} \log f(x) &= \frac{1}{f(x)} \frac{\partial}{\partial x} f(x) \\
\frac{\partial}{\partial \hat{y}_i} \log (1-\hat{y}_i) &= \frac{1}{1-\hat{y}_i} \frac{\partial}{\partial \hat{y}_i} (1-\hat{y}_i)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial x} (x-y) &= \frac{\partial}{\partial x} x - \frac{\partial}{\partial x} y \\
\frac{\partial}{\partial \hat{y}_i} (1-\hat{y}_i) &= \frac{\partial}{\partial \hat{y}_i} 1 - \frac{\partial}{\partial \hat{y}_i} \hat{y}_i \\
&= -1
\end{aligned}$$

Parameters learning $\frac{\partial \mathcal{L}}{\partial \hat{y}_i}$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \hat{y}_i} &= \frac{\partial}{\partial \hat{y}_i} \frac{1}{N} \sum_{i=1}^{N} -y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i) \\
&= \frac{1}{N} \frac{\partial}{\partial \hat{y}_i} \sum_{i=1}^{N} -y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i) \\
&= \frac{1}{N} \sum_{i=1}^{N} -\frac{\partial}{\partial \hat{y}_i} \left(y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i) \right) \\
&= \frac{1}{N} \sum_{i=1}^{N} -\frac{\partial}{\partial \hat{y}_i} y_i \log \hat{y}_i + \frac{\partial}{\partial \hat{y}_i} (1 - y_i) \log (1 - \hat{y}_i) \\
&= \frac{1}{N} \sum_{i=1}^{N} -\frac{y_i}{\hat{y}_i} + \frac{\partial}{\partial \hat{y}_i} (1 - y_i) \log (1 - \hat{y}_i) \\
&= \frac{1}{N} \sum_{i=1}^{N} -\frac{y_i}{\hat{y}_i} - \frac{1 - y_i}{1 - \hat{y}_i}
\end{aligned}
$$

Parameters learning $\frac{\partial \mathcal{L}}{\partial \hat{y}_i}$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \hat{y}_i} &= \frac{\partial}{\partial \hat{y}_i} \frac{1}{N} \sum_{i=1}^{N} -y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i) \\
&= \frac{1}{N} \frac{\partial}{\partial \hat{y}_i} \sum_{i=1}^{N} -y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i) \\
&= \frac{1}{N} \sum_{i=1}^{N} -\frac{\partial}{\partial \hat{y}_i} \left(y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i) \right) \\
&= \frac{1}{N} \sum_{i=1}^{N} -\frac{\partial}{\partial \hat{y}_i} y_i \log \hat{y}_i + \frac{\partial}{\partial \hat{y}_i} (1 - y_i) \log (1 - \hat{y}_i) \\
&= \frac{1}{N} \sum_{i=1}^{N} -\frac{y_i}{\hat{y}_i} + \frac{\partial}{\partial \hat{y}_i} (1 - y_i) \log (1 - \hat{y}_i) \\
&= \frac{1}{N} \sum_{i=1}^{N} -\frac{y_i}{\hat{y}_i} - \frac{1 - y_i}{1 - \hat{y}_i} = \frac{1}{N} \sum_{i=1}^{N} \frac{\hat{y}_i - y_i}{\hat{y}_i (1 - \hat{y}_i)}
\end{aligned}
$$

Parameters learning $\frac{\partial \hat{y}_i}{\partial z_i}$

$$\frac{\partial \hat{y}_i}{\partial z_i} = \frac{\partial}{\partial z_i} \frac{1}{1 + e^{-z_i}}$$

$$\begin{aligned}
\frac{\partial}{\partial x} \frac{f(x)}{g(x)} &= \frac{g(x)\frac{\partial}{\partial x} f(x) - f(x)\frac{\partial}{\partial x} g(x)}{g^2(x)} \\
\frac{\partial}{\partial z_i} \frac{1}{1 + e^{-z_i}} &= \frac{(1 + e^{-z_i})\frac{\partial}{\partial z_i} 1 - 1\frac{\partial}{\partial z_i}(1 + e^{-z_i})}{(1 + e^{-z_i})^2} \\
&= \frac{-1\frac{\partial}{\partial z_i}(1 + e^{-z_i})}{(1 + e^{-z_i})^2} \\
&= -\frac{1}{(1 + e^{-z_i})^2} \frac{\partial}{\partial z_i}(1 + e^{-z_i})
\end{aligned}$$

Parameters learning $\dfrac{\partial \hat{y}_i}{\partial z_i}$

$$\frac{\partial \hat{y}_i}{\partial z_i} = \frac{\partial}{\partial z_i}\frac{1}{1+e^{-z_i}}$$
$$= -\frac{1}{(1+e^{-z_i})^2}\frac{\partial}{\partial z_i}(1+e^{-z_i})$$

$$\boxed{\begin{aligned}
\frac{\partial}{\partial x}(f(x)+g(x)) &= \frac{\partial}{\partial x}f(x) + \frac{\partial}{\partial x}g(x)\\
\frac{\partial}{\partial z_i}(1+e^{-z_i}) &= \frac{\partial}{\partial z_i}1 + \frac{\partial}{\partial z_i}e^{-z_i}\\
&= \frac{\partial}{\partial z_i}e^{-z_i}\\[2em]
\frac{\partial}{\partial x}e^{x} &= e^{x}\frac{\partial}{\partial x}x\\
\frac{\partial}{\partial z_i}e^{-z_i} &= e^{-z_i}\frac{\partial}{\partial z_i}(-z_i)\\
&= e^{-z_i}(-1)\\
&= -e^{-z_i}
\end{aligned}}$$

Parameters learning $\frac{\partial \hat{y}_i}{\partial z_i}$

$$
\begin{aligned}
\frac{\partial \hat{y}_i}{\partial z_i} &= \frac{\partial}{\partial z_i} \frac{1}{1+e^{-z_i}} \\
&= -\frac{1}{(1+e^{-z_i})^2} \frac{\partial}{\partial z_i}(1+e^{-z_i}) \\
&= \frac{e^{-z_i}}{(1+e^{-z_i})^2}
\end{aligned}
$$

Parameters learning $\frac{\partial \hat{y}_i}{\partial z_i}$

$$
\begin{aligned}
\frac{\partial \hat{y}_i}{\partial z_i} &= \frac{\partial}{\partial z_i} \frac{1}{1+e^{-z_i}} \\
&= -\frac{1}{(1+e^{-z_i})^2} \frac{\partial}{\partial z_i}(1+e^{-z_i}) \\
&= \frac{e^{-z_i}}{(1+e^{-z_i})^2} \\
&= \frac{1}{1+e^{-z_i}} \frac{e^{-z_i}}{1+e^{-z_i}}
\end{aligned}
$$

Parameters learning $\frac{\partial \hat{y}_i}{\partial z_i}$

$$
\begin{aligned}
\frac{\partial \hat{y}_i}{\partial z_i} &= \frac{\partial}{\partial z_i} \frac{1}{1 + e^{-z_i}} \\
&= -\frac{1}{(1 + e^{-z_i})^2} \frac{\partial}{\partial z_i} (1 + e^{-z_i}) \\
&= \frac{e^{-z_i}}{(1 + e^{-z_i})^2} \\
&= \frac{1}{1 + e^{-z_i}} \frac{e^{-z_i}}{1 + e^{-z_i}} \\
&= \frac{1}{1 + e^{-z_i}} \frac{1 + e^{-z_i} - 1}{1 + e^{-z_i}} = \frac{1}{1 + e^{-z_i}} \frac{(1 + e^{-z_i}) - 1}{1 + e^{-z_i}}
\end{aligned}
$$

Parameters learning $\dfrac{\partial \hat{y}_i}{\partial z_i}$

$$
\begin{aligned}
\frac{\partial \hat{y}_i}{\partial z_i} &= \frac{\partial}{\partial z_i}\frac{1}{1+e^{-z_i}} \\
&= -\frac{1}{(1+e^{-z_i})^2}\frac{\partial}{\partial z_i}(1+e^{-z_i}) \\
&= \frac{e^{-z_i}}{(1+e^{-z_i})^2} \\
&= \frac{1}{1+e^{-z_i}}\frac{e^{-z_i}}{1+e^{-z_i}} \\
&= \frac{1}{1+e^{-z_i}}\frac{1+e^{-z_i}-1}{1+e^{-z_i}} = \frac{1}{1+e^{-z_i}}\frac{(1+e^{-z_i})-1}{1+e^{-z_i}} \\
&= \frac{1}{1+e^{-z_i}}\left(\frac{1+e^{-z_i}}{1+e^{-z_i}} - \frac{1}{1+e^{-z_i}}\right)
\end{aligned}
$$

Parameters learning $\frac{\partial \hat{y}_i}{\partial z_i}$

$$
\begin{aligned}
\frac{\partial \hat{y}_i}{\partial z_i} &= \frac{\partial}{\partial z_i} \frac{1}{1+e^{-z_i}} \\
&= -\frac{1}{(1+e^{-z_i})^2} \frac{\partial}{\partial z_i}(1+e^{-z_i}) \\
&= \frac{e^{-z_i}}{(1+e^{-z_i})^2} \\
&= \frac{1}{1+e^{-z_i}} \frac{e^{-z_i}}{1+e^{-z_i}} \\
&= \frac{1}{1+e^{-z_i}} \frac{1+e^{-z_i}-1}{1+e^{-z_i}} = \frac{1}{1+e^{-z_i}} \frac{(1+e^{-z_i})-1}{1+e^{-z_i}} \\
&= \frac{1}{1+e^{-z_i}} \left(\frac{1+e^{-z_i}}{1+e^{-z_i}} - \frac{1}{1+e^{-z_i}} \right) \\
&= \hat{y}_i \, (1-\hat{y}_i)
\end{aligned}
$$

Parameters learning $\frac{\partial z_i}{\partial w_j}$

$$\frac{\partial z_i}{\partial w_j} = \frac{\partial}{\partial w_j}\left(w_0 + \sum_{j=1}^{n} x_{i,j} w_j\right)$$

Parameters learning $\frac{\partial z_i}{\partial w_j}$

$$\frac{\partial z_i}{\partial w_j} = \frac{\partial}{\partial w_j}\left(w_0 + \sum_{j=1}^{n} x_{i,j} w_j\right)$$

$$= \frac{\partial}{\partial w_j}\left(w_0 + x_{i,1} w_1 + \ldots + x_{i,n} w_n\right)$$

Parameters learning $\frac{\partial z_i}{\partial w_j}$

$$\frac{\partial z_i}{\partial w_j} = \frac{\partial}{\partial w_j}\left(w_0 + \sum_{j=1}^{n} x_{i,j} w_j\right)$$

$$= \frac{\partial}{\partial w_j}\left(w_0 + x_{i,1}w_1 + \ldots + x_{i,n}w_n\right)$$

$$= \frac{\partial}{\partial w_j}w_0 + \frac{\partial}{\partial w_j}x_{i,1}w_1 + \ldots + \frac{\partial}{\partial w_j}x_{i,n}w_n$$

Parameters learning $\frac{\partial z_i}{\partial w_j}$

$$\frac{\partial z_i}{\partial w_j} = \frac{\partial}{\partial w_j}\left(w_0 + \sum_{j=1}^{n} x_{i,j} w_j\right)$$

$$= \frac{\partial}{\partial w_j}\left(w_0 + x_{i,1}w_1 + \ldots + x_{i,n}w_n\right)$$

$$= \frac{\partial}{\partial w_j} w_0 + \frac{\partial}{\partial w_j} x_{i,1}w_1 + \ldots + \frac{\partial}{\partial w_j} x_{i,n}w_n$$

$$= \frac{\partial}{\partial w_j} x_{i,j}w_j$$

Parameters learning $\frac{\partial z_i}{\partial w_j}$

$$
\begin{aligned}
\frac{\partial z_i}{\partial w_j} &= \frac{\partial}{\partial w_j}\left(w_0 + \sum_{j=1}^{n} x_{i,j} w_j\right) \\
&= \frac{\partial}{\partial w_j}\left(w_0 + x_{i,1}w_1 + \ldots + x_{i,n}w_n\right) \\
&= \frac{\partial}{\partial w_j} w_0 + \frac{\partial}{\partial w_j} x_{i,1}w_1 + \ldots + \frac{\partial}{\partial w_j} x_{i,n}w_n \\
&= \frac{\partial}{\partial w_j} x_{i,j}w_j \\
&= x_{i,j}\frac{\partial}{\partial w_j} w_j + w_j \frac{\partial}{\partial w_j} x_{i,j}
\end{aligned}
$$

Parameters learning $\frac{\partial z_i}{\partial w_j}$

$$
\begin{aligned}
\frac{\partial z_i}{\partial w_j} &= \frac{\partial}{\partial w_j}\left(w_0 + \sum_{j=1}^{n} x_{i,j} w_j\right) \\
&= \frac{\partial}{\partial w_j}\left(w_0 + x_{i,1}w_1 + \ldots + x_{i,n}w_n\right) \\
&= \frac{\partial}{\partial w_j} w_0 + \frac{\partial}{\partial w_j} x_{i,1}w_1 + \ldots + \frac{\partial}{\partial w_j} x_{i,n}w_n \\
&= \frac{\partial}{\partial w_j} x_{i,j} w_j \\
&= x_{i,j}\frac{\partial}{\partial w_j} w_j + w_j \frac{\partial}{\partial w_j} x_{i,j} \\
&= x_{i,j}
\end{aligned}
$$

Parameters learning

$$\frac{\partial \mathcal{L}}{\partial \hat{y}_i} = \frac{1}{N}\sum_{i=1}^{N} \frac{\hat{y}_i - y_i}{\hat{y}_i(1-\hat{y}_i)}$$

$$\frac{\partial \hat{y}_i}{\partial z_i} = \hat{y}_i\left(1-\hat{y}_i\right)$$

$$\frac{\partial z_i}{\partial w_j} = x_{i,j}$$

$$\boxed{\begin{aligned}\frac{\partial \mathcal{L}}{\partial w_j} &= \frac{\partial \mathcal{L}}{\partial \hat{y}_i}\frac{\partial \hat{y}_i}{\partial z_i}\frac{\partial z_i}{\partial w_j} \\ &= \frac{1}{N}\sum_{i=1}^{N} \frac{\hat{y}_i - y_i}{\hat{y}_i(1-\hat{y}_i)}\hat{y}_i\left(1-\hat{y}_i\right)x_{i,j} \\ &= \frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)x_{i,j}\end{aligned}}$$

Parameters learning

$$w_j \leftarrow w_j - \eta \frac{\partial \mathcal{L}}{\partial w_j}$$

$$\frac{\partial \mathcal{L}}{\partial w_j} = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i) x_{i,j}$$

$$w_j \leftarrow w_j - \frac{\eta}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i) x_{i,j}$$

Parameters learning

Given $\mathbf{D} = \{(\boldsymbol{x}_1, y_1), \dots, (\boldsymbol{x}_N, y_N)\}$,
where $\boldsymbol{x}_i = \{x_{i,1}, \dots, x_{i,n}\}$;
Randomly generate parameters $w_0, \dots, w_n$;
Set the learning rate $\eta \in (0, 1]$;
Set the number of epochs N_{epoch};
$t \leftarrow 0$;
while $t \leq N_{epoch}$ **do**
  $t \leftarrow t + 1$;
  $\Delta w_j = 0$;
  foreach $(\boldsymbol{x}_i, y_i) \in \mathbf{D}$ **do**
    $\hat{y}_i = \frac{1}{1+e^{-(w_0+\sum_{j=1}^{n} w_j x_{i,j})}}$;
    $\Delta w_j \leftarrow \Delta w_j + (\hat{y}_i - y_i)\, x_{i,j}$;
  end
  $w_j \leftarrow w_j - \frac{\eta}{N} \Delta w_j$;
end

Example - epoch=3, η=0.01, round 1

w0	w1	w2
0.0	0.1	0.2

Training dataset

x1	x2	y
0	0	0
0	1	1
1	0	1
1	1	1

$$\frac{1}{1+e^{-(1\times0.0+0\times0.1+0\times0.2)}} = 0.50, \hat{y}-y = 0.50$$

$$\frac{1}{1+e^{-(1\times0.0+0\times0.1+1\times0.2)}} = 0.55, \hat{y}-y = -0.45$$

$$\frac{1}{1+e^{-(1\times0.0+1\times0.1+0\times0.2)}} = 0.52, \hat{y}-y = -0.48$$

$$\frac{1}{1+e^{-(1\times0.0+1\times0.1+1\times0.2)}} = 0.57, \hat{y}-y = -0.43$$

$$\begin{aligned} w_0 &\leftarrow 0.0 - \frac{0.01}{4} \times (0.50 - 0.45 - 0.48 - 0.43) \\ &\leftarrow 0.002 \\ w_1 &\leftarrow 0.1 - \frac{0.01}{4} \times (0.50 \times 0 - 0.45 \times 0 - 0.48 \times 1 - 0.43 \times 1) \\ &\leftarrow 0.102 \\ w_2 &\leftarrow 0.2 - \frac{0.01}{4} \times (0.50 \times 0 - 0.45 \times 1 - 0.48 \times 0 - 0.43 \times 1) \\ &\leftarrow 0.202 \end{aligned}$$

Example - epoch=3, η=0.01, round 2

w0	w1	w2
0.002	0.102	0.202

$$\frac{1}{1+e^{-(1\times0.002+0\times0.102+0\times0.202)}} = 0.5, \hat{y}-y = 0.50$$

$$\frac{1}{1+e^{-(1\times0.002+0\times0.102+1\times0.202)}} = 0.55, \hat{y}-y = -0.45$$

$$\frac{1}{1+e^{-(1\times0.002+1\times0.102+0\times0.202)}} = 0.53, \hat{y}-y = -0.47$$

$$\frac{1}{1+e^{-(1\times0.002+1\times0.102+1\times0.202)}} = 0.58, \hat{y}-y = -0.42$$

Training dataset

x1	x2	y
0	0	0
0	1	1
1	0	1
1	1	1

$$w_0 \leftarrow 0.002 - \frac{0.01}{4} \times (0.5 - 0.45 - 0.47 - 0.42)$$

$$\leftarrow 0.004$$

$$w_1 \leftarrow 0.102 - \frac{0.01}{4} \times (0.50\times 0 - 0.45\times 0 - 0.47\times 1 - 0.42\times 1)$$

$$\leftarrow 0.104$$

$$w_2 \leftarrow 0.202 - \frac{0.01}{4} \times (0.50\times 0 - 0.45\times 1 - 0.47\times 0 - 0.42\times 1)$$

$$\leftarrow 0.204$$

Example - epoch=3, η=0.01, round 3

w0	w1	w2
0.004	0.104	0.204

Training dataset

x1	x2	y
0	0	0
0	1	1
1	0	1
1	1	1

$$\frac{1}{1+e^{-(1\times 0.004+0\times 0.104+0\times 0.204)}} = 0.5, \hat{y}-y = 0.50$$

$$\frac{1}{1+e^{-(1\times 0.004+0\times 0.104+1\times 0.204)}} = 0.55, \hat{y}-y = -0.45$$

$$\frac{1}{1+e^{-(1\times 0.004+1\times 0.104+0\times 0.204)}} = 0.53, \hat{y}-y = -0.47$$

$$\frac{1}{1+e^{-(1\times 0.004+1\times 0.104+1\times 0.204)}} = 0.58, \hat{y}-y = -0.42$$

$$w_0 \leftarrow 0.004 - \frac{0.01}{4} \times (0.5 - 0.45 - 0.47 - 0.42)$$

$$\leftarrow 0.006$$

$$w_1 \leftarrow 0.104 - \frac{0.01}{4} \times (0.50 \times 0 - 0.45 \times 0 - 0.47 \times 1 - 0.42 \times 1)$$

$$\leftarrow 0.106$$

$$w_2 \leftarrow 0.204 - \frac{0.01}{4} \times (0.50 \times 0 - 0.45 \times 1 - 0.47 \times 0 - 0.42 \times 1)$$

$$\leftarrow 0.206$$

Workshop 1

Calculate the learning parameters for logistic regression at the 4th epoch.

Workshop 2

Calculate the learning parameters for logistic regression using the dataset below, where all initialized learning parameters are set to 0, the learning rate is 0.01, and the number of epochs is 1.

Dataset

x1	x2	y
0	0	0
0	1	0
1	0	0
1	1	1