$$J(\theta) = -\frac{1}{N}\sum_{i=1}^{N}\Big(y_i \log(h_\theta(x_i)) + (1-y_i)\log(1-h_\theta(x_i))\Big)$$

① → $$\frac{\partial J(\theta)}{\partial(\theta_j)} = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i * \frac{1}{h_\theta(x_i)} * \frac{\partial h_\theta(x_i)}{\partial(\theta_j)} + (1-y_i) * \frac{1}{(1-h_\theta(x_i))} \cdot \frac{\partial(1-h_\theta(x_i))}{\partial(\theta_j)}\right)$$

$$+ (1-y_i) * \frac{1}{(1-h_\theta(x_i))} \cdot \frac{(-1)\partial(h_\theta(x_i))}{\partial(\theta_j)}$$

$$\boxed{\underline{\text{Note}} \quad \frac{\partial\sigma(x)}{\partial(x)} = \sigma(x)(1-\sigma(x))}$$

② → $$\frac{\partial J(\theta)}{\partial(\theta_j)} = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i * \frac{1}{h_\theta(x_i)} * \frac{\partial\sigma(\theta^T x)}{\partial(\theta_j)} + (1-y_i) * \frac{1}{(1-h_\theta(x_i))} \cdot (-1)\frac{\partial\sigma(\theta^T x)}{\partial(\theta_j)}\right)$$

③ → $$= -\frac{1}{N}\sum_{i=1}^{N}\left(y_i \cdot \frac{1}{h_\theta(x_i)} \cdot \sigma(\theta^T x) * (1-\sigma(\theta^T x)) * x_{ij} + (1-y_i) \cdot \frac{1}{(1-h_\theta(x_i))} \cdot (-1)\,\sigma(\theta^T x) \cdot (1-\sigma(\theta^T x)) \cdot x_{ij}\right)$$

$$\frac{\partial \sigma(\theta^T x)}{\partial(\theta_j)} = \frac{\partial \sigma(\theta^T x)}{\partial(\theta^T x)} \times \boxed{\frac{\partial(\theta^T x)}{\partial \theta_j}}$$

$$= \sigma(\theta^T x)\cdot(1-\sigma(\theta^T x))\cdot x_{i,j}$$

$$= h_\theta(x_i)\cdot(1-h_\theta(x_i))\, x_{i,j}$$

$$\therefore \quad h_\theta(x_i) \equiv \sigma(\theta^T x)$$

4 $$= -\frac{1}{N}\sum_{i=1}^{N}\left(y_i \cdot \cancel{\frac{1}{h_\theta(x_i)}} \cdot \cancel{\sigma(\theta^T x)} \times (1-\sigma(\theta^T x)) \times x_{ij} + (1-y_i)\cdot \cancel{\frac{1}{(1-h_\theta(x_i))}} \cdot (-1)\,\sigma(\theta^T x)\cdot \cancel{(1-\sigma(\theta^T x))}\cdot x_{ij} \right)$$

5 $$= -\frac{1}{N}\sum_{i=1}^{N}\Big(y_i\,(1-h_\theta(x_i))\, x_{ij} + (1-y_i)\,(-1)\,h_\theta(x_i)\, x_{i,j}$$

$$\frac{\partial J(\theta)}{\partial \theta_j} = -\frac{1}{N}\sum_{i=1}^{N}\Big(y_i\,(1-h_\theta(x_i))\,x_{ij} + (1-y_i)(-1)\,h_\theta(x_i)\,x_{ij}$$

$$= -\frac{1}{N}\sum_{i=1}^{N}\Big(y_i x_{ij} - \cancel{y_i h_\theta(x_i) x_{ij}} + \cancel{y_i h_\theta(x_i) x_{ij}} - h_\theta(x_i) x_{ij}$$

$$= -\frac{1}{N}\sum_{i=1}^{N}\big(y_i x_{ij} - h_\theta(x_i) x_{ij} \big)$$

$$= -\frac{1}{N}\sum_{i=1}^{N}\big(y_i - h_\theta(x_i) \big) x_{ij}$$

$$= \boxed{\frac{1}{N}\sum_{i=1}^{N}\big(h_\theta(x_i) - y_i \big) x_{ij}}$$

จะเป็นฟอร์มเดียวกับ GD ของ MSE

	Regression	Logistic
Cost function objective function	MSE	NLL
Evaluation function (on test set)	MSE R^2 RMSE MAE MPE	y $\hat{y}$ 1 1 0 1 1 1 1 1 Accuracy $\frac{TP+TN}{Total}$ Confusion Matrix F1 score, Precision, Recall

Appendix

1. **Odds** → the ratio of the probability that the event will happen to the probability that the event will not happen.

$= \frac{P(event)}{1 - P(event)}$ Ex. $\frac{0.8}{0.2} = 4$, $\frac{0.9}{0.1} = 9$, $\frac{0.5}{0.5} = 1$, $\frac{0.2}{0.8} = 0.25$

2. **log odds** →

$\ln(4) = 1.386$, $\ln(9) = 2.197$, $\ln(1) = 0$, $\ln(0.25) = -1.386$

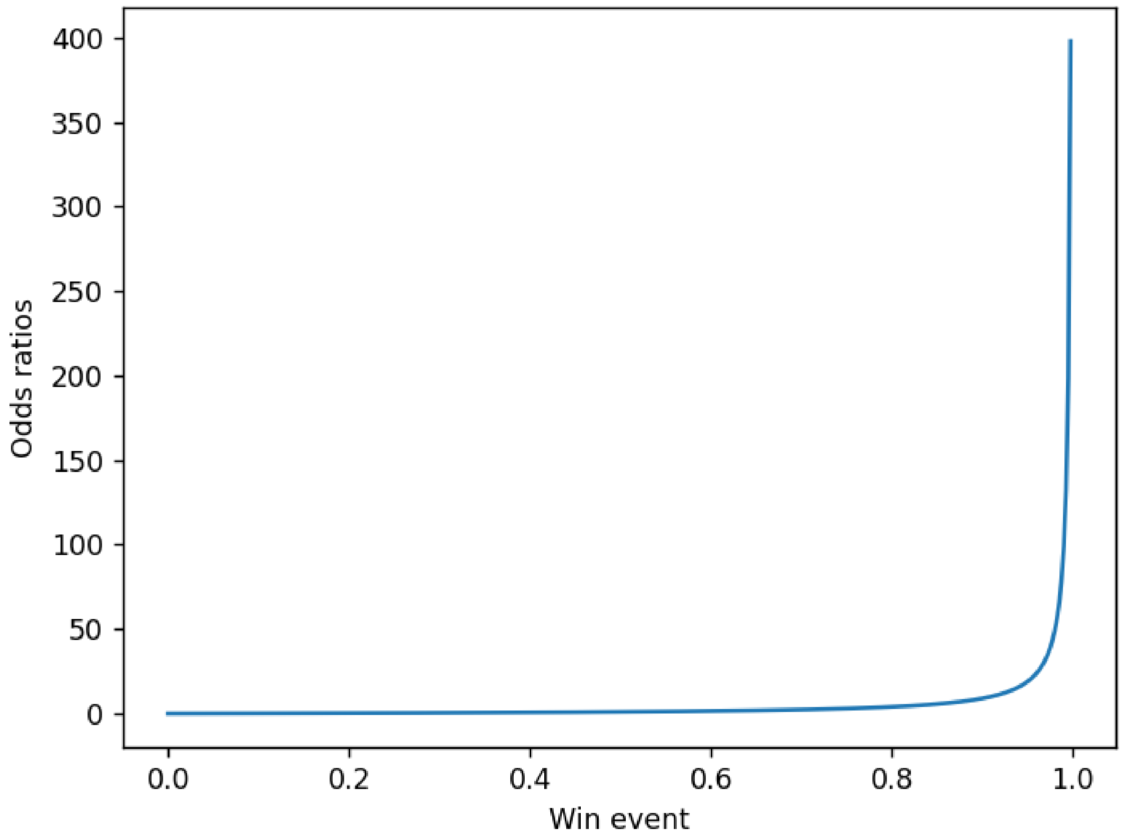
log odd ratio also called "logit function"

input is $(0, 1)$

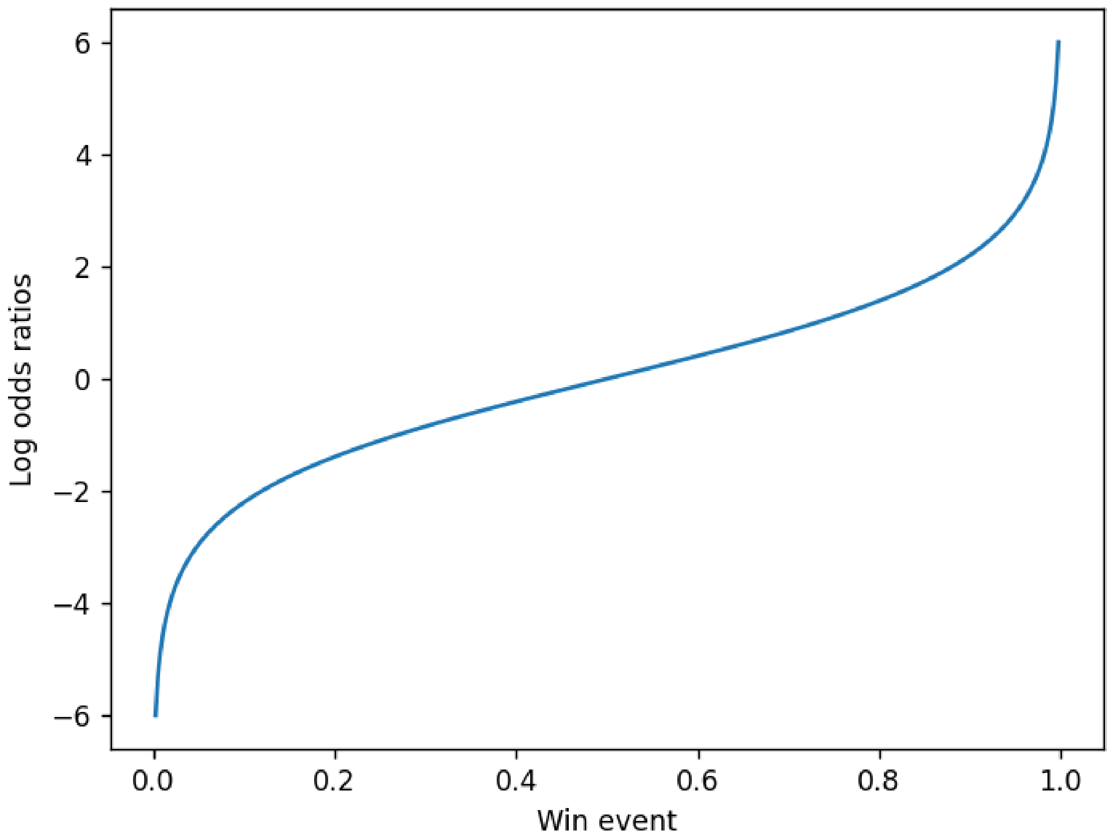
output is $(-\infty, +\infty)$

Every probability can be easily converted to **log odds**, by finding the odds ratio and taking the logarithm.

```
import matplotlib.pyplot as plt
import numpy as np

xval = np.linspace(0,1,400)[0:399]
odds_ratio = xval/(1-xval)

plt.plot(xval,odds_ratio)
plt.xlabel("Win event")
plt.ylabel("Odds ratios")
plt.show()
```



```
import matplotlib.pyplot as plt
import numpy as np

xval = np.linspace(0,1,400)[0:399]
odds_ratio = xval/(1-xval)

plt.plot(xval,np.log(odds_ratio))
plt.xlabel("Win event")
plt.ylabel("Log odds ratios")
plt.show()
```



Relationship between Linear combination and Probabi-

$$P(y=1|x) = \frac{e^{\theta^T x}}{1+e^{\theta^T x}}$$

$$P(y=0|x) = 1 - \frac{e^{\theta^T x}}{1+e^{\theta^T x}} = \frac{1}{1+e^{\theta^T x}}$$

odds ratio

$$\frac{P}{1-P} = \frac{\frac{e^{\theta^T x}}{1+e^{\theta^T x}}}{\frac{1}{1+e^{\theta^T x}}} = e^{\theta^T x}$$

Log odds ratio/logit

$$\log\left(\frac{P}{1-P}\right) = \log\left(e^{\theta^T x}\right) = \theta^T x = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots$$

$\{0, 1\}$ and p be the probability of y to be 1, $p = P(Y = 1)$. Let $x_1, \cdots, x_k$ be a set of predictor variables. Then the logistic regression of
estimates parameter values for $\beta_0, \beta_1, \cdots, \beta_k$ via maximum likelihood method of the following equation

$$logit(p) = log(\frac{p}{1-p}) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k.$$

Exponentiate and take the multiplicative inverse of both sides,

$$\frac{1-p}{p} = \frac{1}{exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)}.$$

Partial out the fraction on the left-hand side of the equation and add one to both sides,

$$\frac{1}{p} = 1 + \frac{1}{exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)}.$$

Change 1 to a common denominator,

$$\frac{1}{p} = \frac{exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k) + 1}{exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)}.$$

Finally, take the multiplicative inverse again to obtain the formula for the probability $P(Y = 1)$,

$$p = \frac{exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)}{1 + exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)}.$$

https://stats.oarc.ucla.edu/other/mult-pkg/faq/general/faq-how-do-i-interpret-odds-ratios-in-logistic-regression/