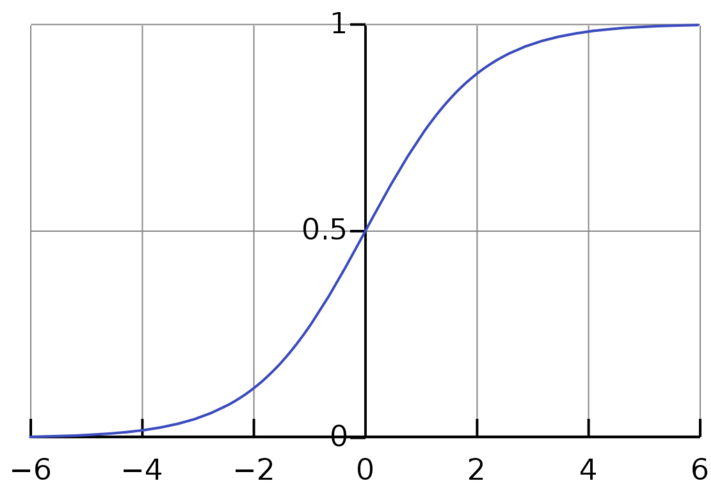
2. Logistic Regression

Logistic regression is a statistical method used for binary classification tasks. It's an extension of linear regression, where the dependent variable is categorical (usually binary) rather than continuous.

The logistic regression model predicts the probability that a given input belongs to a certain category. It accomplishes this by applying the logistic function (also known as the sigmoid function) to a linear combination of the input features. The logistic function ensures that the output of the regression model lies between 0 and 1, which can be interpreted as probabilities.

Sigmoid function:

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



Logistic regression:

$$P(y=1|\mathbf{x}) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_n x_n)}}$$

- $P(y=1 \mid \mathbf{x})$ is the probability that the dependent variable y is 1 given the input features **x**.
- e is the base of the natural logarithm.
- $\beta_0, \beta_1, \ldots, \beta_n$ are the coefficients of the model.
- $x_1, x_2, \ldots, x_n$ are the input features.

Binary cross-entropy

The loss function used in logistic regression is typically the binary cross-entropy loss function, also known as the log loss or logistic loss. It measures the difference between the predicted probabilities of the model and the actual binary labels in the training data.

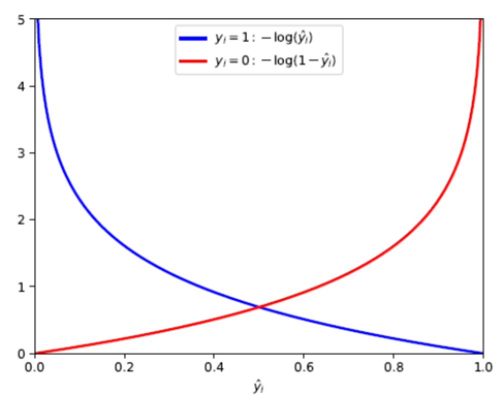
$$L(y, \hat{y}) = -\frac{1}{N} \sum_{i=1}^{N} \left(y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right)$$

Where:

- *L*(*y*,*y*^) is the binary cross-entropy loss.
- *N* is the number of samples in the training data.
- *yi* is the actual binary label (0 or 1) for the *i*th sample.
- *y^i* is the predicted probability that the *i*th sample belongs to the positive class (i.e., *yi*=1).

$$\begin{aligned} \mathcal{L}(\hat{y}_i, y_i) &= -y_i \log \hat{y}_i - (1 - y_i) \log(1 - \hat{y}_i) \\ &= \begin{cases} -\log \hat{y}_i & \text{if } y_i = 1 \\ -\log(1 - \hat{y}_i) & \text{if } y_i = 0 \end{cases} \end{aligned}$$

- For $y_i = 1$ the loss $\mathcal{L}$ is minimized if $\hat{y}_i = 1$
- For $y_i = 0$ the loss $\mathcal{L}$ is minimized if $\hat{y}_i = 0$
- Thus, $\mathcal{L}$ is minimal if $\hat{y}_i = y_i$
- Can be extended to > 2 classes



Linear Regression	Logistic Regression
Dependent variable is continuous	Dependent variable is categorical (usually binary)
Predicts the expected value of dependent variable	Predicts the probability of occurrence of an event
Output value can be any number	Output value lies between 0 and 1 [0,1]
Residuals are assumed to be normally distributed	Doesn't make this assumption
Uses mean squared error as a loss function	Uses binary cross-entropy (log loss) as a loss function

If you have the mean and standard deviation of a dataset, you can use them to plot any value within that dataset. This concept is closely related to the idea of the normal distribution (also known as Gaussian distribution).

Here's how it works:

1. **Mean (Average)**: The mean represents the central tendency of the data. It's the sum of all the values in the dataset divided by the number of values. It gives you a single value that is representative of the dataset.
2. **Standard Deviation**: The standard deviation measures the dispersion or spread of the data points around the mean. A low standard deviation indicates that the data points tend to be close to the mean, while a high standard deviation indicates that the data points are spread out over a wider range of values.

When you have the mean and standard deviation of a dataset, you can use them to understand the distribution of the data and make predictions about where new data points are likely to fall.

:= assigment