logistics regression

linear classification: find the hyperplane

$$\theta^T x + b = 0$$

threshold.

$$\begin{cases} \theta^T x \geq -b \rightarrow y = 0 \\ \theta^T x < -b \rightarrow y = 1 \end{cases}$$

$$\downarrow$$

predict result.

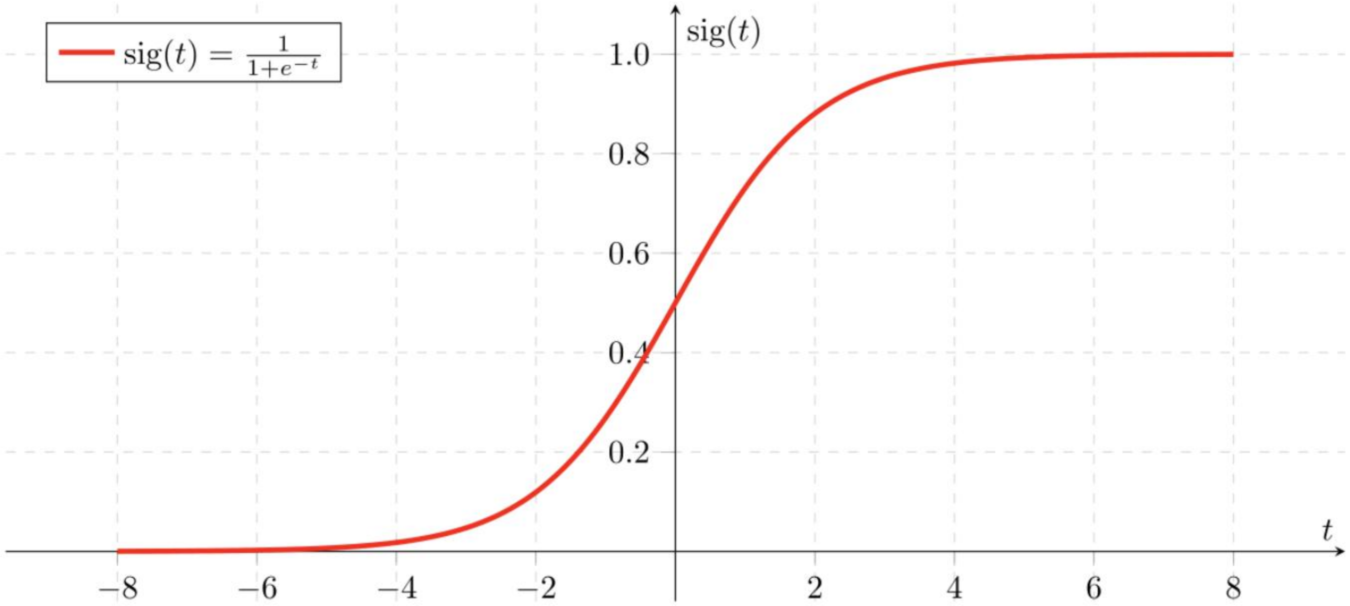
$$\hookrightarrow h_\theta(x) = g\left(\theta^T x\right) = \frac{1}{1+e^{-\theta^T x}} = \frac{1}{1+e^{-z}}$$

$$\begin{cases} p(Y=0|x) = \dfrac{1}{1+e^{-\theta^T x}} \\ p(Y=1|x) = \dfrac{e^{-\theta^T x}}{1+e^{-\theta^T x}} \end{cases} \rightarrow \text{probability distribution}$$

logit function

$$\text{map } \theta^T x(z) : -\infty \sim \infty \text{ to } 0 \sim 1$$

If ‘Z’ goes to infinity, Y(predicted) will become 1 and if ‘Z’ goes to negative infinity, Y(predicted) will become 0.



Binary Logistic Regression

We view each observasions y_i as an independent sample from a Bernoulli distribution $Y_i \sim Bern(p_i)$, (technically we mean $\hat{Y}_i|\mathbf{x}_i, \mathbf{w}$), where p_i is a function of $\mathbf{x}_i$.

We need a model for the dependency of p_i on $\mathbf{x}_i$. We have to enforce that p_i is a transformation of $\mathbf{x}_i$ that results in a number from 0 to 1 (ie. a valid probability). Hence p_i cannot be, say, linear in x_i. One way to do achieve the 0-1 normalization is by using the sigmoid function.

$$p_i = s(\mathbf{w}^T \mathbf{x}_i) = \frac{1}{1 + e^{-\mathbf{w}^T \mathbf{x}_i}}$$

$$
\begin{aligned}
\hat{\mathbf{w}} &= \arg\max_{\mathbf{w}} P(y_1, \cdots, y_n | \mathbf{x}_1, \cdots, \mathbf{x}_n, \mathbf{w}) \\
&= \arg\max_{\mathbf{w}} \prod_{i=1}^{n} P(y_i | \mathbf{x}_i, \mathbf{w}) \quad \text{iid.} \\
&\text{(max-log-likelihood).} \\
&= \arg\max_{\mathbf{w}} \ln[\prod_{i=1}^{n} p_i^{y_i} (1 - p_i)^{(1-y_i)}] \\
&= \arg\max_{\mathbf{w}} \sum_{i=1}^{n} y_i \ln p_i + (1 - y_i) \ln(1 - p_i) \\
&= \arg\min_{\mathbf{w}} \underbrace{- \sum_{i=1}^{n} y_i \ln p_i + (1 - y_i) \ln(1 - p_i)}_{\text{loss function.}}
\end{aligned}
$$

Binary Logistic Regression

$$
\begin{aligned}
L(\mathbf{w}) &= -\sum_{i=1}^{n} y_i \ln p_i + (1 - y_i) \ln(1 - p_i) \\
\nabla_{\mathbf{w}} L(\mathbf{w}) &= \nabla_{\mathbf{w}} \left(-\sum_{i=1}^{n} y_i \ln p_i + (1 - y_i) \ln(1 - p_i) \right) \\
&= -\sum_{i=1}^{n} y_i \nabla_{\mathbf{w}} \ln p_i + (1 - y_i) \nabla_{\mathbf{w}} \ln(1 - p_i) \\
&= -\sum_{i=1}^{n} \frac{y_i}{p_i} \nabla_{\mathbf{w}} p_i - \frac{1 - y_i}{1 - p_i} \nabla_{\mathbf{w}} p_i
\end{aligned}
$$

Binary Logistic Regression

$$
\begin{aligned}
p_i &= s(\mathbf{w}^T x_i) = \frac{1}{1 + e^{-\mathbf{w}^T x_i}} \\
\nabla_z s(z) &= \nabla_z (1 + e^{-z})^{-1} \\
&= \frac{e^{-z}}{(1 + e^{-z})^2} \\
&= s(z)(1 - s(z)) \\
\nabla_{\mathbf{w}} L(\mathbf{w}) &= -\sum_{i=1}^{n} \frac{y_i}{p_i} \nabla_{\mathbf{w}} p_i - \frac{1 - y_i}{1 - p_i} \nabla_{\mathbf{w}} p_i \\
&= -\sum_{i=1}^{n} \left(\frac{y_i}{p_i} - \frac{1 - y_i}{1 - p_i} \right) p_i (1 - p_i) x_i \\
&= -\sum_{i=1}^{n} (y_i - p_i) x_i
\end{aligned}
$$