CSC 412 Machine Learning and Knowledge Discovery

Logistic Regression

1 Model Prediction

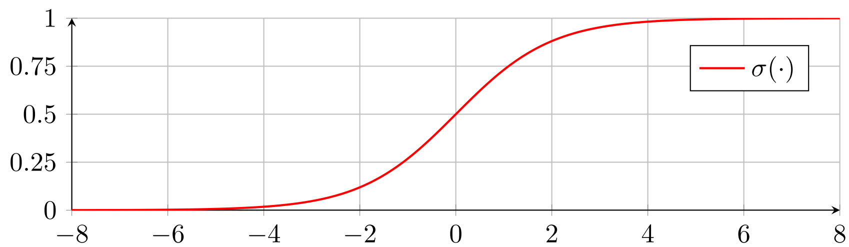
1.1 Vectorized Form

$$\boxed{\hat{p} = h_{\boldsymbol{\theta}}(\boldsymbol{x}) = \sigma(z) = \sigma(\boldsymbol{\theta}^\intercal \boldsymbol{x})}$$

$$\hat{y} = \begin{cases} 0 & \text{if } \hat{p} < 0.5 \\ 1 & \text{if } \hat{p} \geqslant 0.5 \end{cases}$$

1.2 Sigmoid Function

$$\boxed{\sigma(z) = \frac{1}{1 + e^{-z}}}$$



2 Cost Function

$$\left.\begin{array}{l} \mathsf{P}(\, y = 1 \mid \boldsymbol{x}; \boldsymbol{\theta} \,) = h_{\boldsymbol{\theta}}(\boldsymbol{x}) \\ \mathsf{P}(\, y = 0 \mid \boldsymbol{x}; \boldsymbol{\theta} \,) = 1 - h_{\boldsymbol{\theta}}(\boldsymbol{x}) \end{array}\right\} \Rightarrow \mathsf{P}(\, y \mid \boldsymbol{x}; \boldsymbol{\theta} \,) = \big(h_{\boldsymbol{\theta}}(\boldsymbol{x})\big)^{y} \cdot \big(1 - h_{\boldsymbol{\theta}}(\boldsymbol{x})\big)^{1-y}$$

$$= \hat{p}^{y} \cdot (1 - \hat{p})^{1-y}$$

So, **Likelihood**

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) &= \mathsf{P}(\, \boldsymbol{y} \mid \boldsymbol{X}; \boldsymbol{\theta} \,) \\ &= \prod_{i=1}^{m} \mathsf{P}(\, y^{(i)} \mid \boldsymbol{x}^{(i)}; \boldsymbol{\theta} \,) \\ &= \prod_{i=1}^{m} \Big[\big(h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)})\big)^{y^{(i)}} \cdot \big(1 - h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)})\big)^{1-y^{(i)}} \Big] \\ &= \prod_{i=1}^{m} \Big[\big(\hat{p}^{(i)}\big)^{y^{(i)}} \cdot \big(1 - \hat{p}^{(i)}\big)^{1-y^{(i)}} \Big] \end{aligned}$$

Maximizing the Likelihood is equivalent to minimizing the **Negative Log-Likelihood**:

$$\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}) \Leftrightarrow \min_{\boldsymbol{\theta}} -\mathcal{LL}(\boldsymbol{\theta})$$

where

$$\begin{aligned}
-\mathcal{LL}(\boldsymbol{\theta}) &= -\log \mathcal{L}(\boldsymbol{\theta}) \\
&= -\log \prod_{i=1}^{m} \Big[\left(\hat{p}^{(i)}\right)^{y^{(i)}} \cdot \left(1 - \hat{p}^{(i)}\right)^{1-y^{(i)}} \Big] \\
&= -\sum_{i=1}^{m} \Big[\log\left(\left(\hat{p}^{(i)}\right)^{y^{(i)}}\right) + \log\left(\left(1 - \hat{p}^{(i)}\right)^{1-y^{(i)}}\right) \Big] \\
&= -\sum_{i=1}^{m} \left[y^{(i)} \cdot \log\left(\hat{p}^{(i)}\right) + (1 - y^{(i)}) \cdot \log\left(1 - \hat{p}^{(i)}\right) \right]
\end{aligned}$$

So, let

$$\boxed{J(\boldsymbol{\theta}) = -\frac{1}{m}\mathcal{LL}(\boldsymbol{\theta}) = -\frac{1}{m}\sum_{i=1}^{m} \Big[y^{(i)} \log\left(\hat{p}^{(i)}\right) + (1 - y^{(i)}) \log\left(1 - \hat{p}^{(i)}\right) \Big]}$$