Logistic Regression

1 Basic

† $y = \sigma(x) = \frac{1}{1+e^{-x}} = (1+e^{-x})^{-1} \to$

$$\frac{dy}{dx} = (-1)\cdot(1+e^{-x})^{-2}\cdot\frac{d(1+e^{-x})}{dx} = (-1)\cdot(1+e^{-x})^{-2}\cdot e^{-x}\cdot\frac{d(-x)}{dx} = (-1)\cdot(1+e^{-x})^{-2}\cdot e^{-x}\cdot(-1)$$

$$= (1+e^{-x})^{-2}\cdot e^{-x} = \frac{1}{1+e^{-x}}\cdot\frac{e^{-x}}{1+e^{-x}} = \frac{1}{1+e^{-x}}\cdot\left(1-\frac{1}{1+e^{-x}}\right) = \sigma(x)(1-\sigma(x))$$

2 Case : Squared Error Function

2.1 Partial Derivative

† $y = \sigma(w) = \frac{1}{1+e^{-wx}} = (1+e^{-wx})^{-1} \to$

$$\frac{\partial y}{\partial w} = (-1)\cdot(1+e^{-wx})^{-2}\cdot\frac{\partial(1+e^{-wx})}{\partial w} = (-1)\cdot(1+e^{-wx})^{-2}\cdot e^{-wx}\cdot\frac{\partial(-wx)}{\partial w} = (-1)\cdot(1+e^{-wx})^{-2}\cdot e^{-wx}\cdot(-x)$$

$$= (1+e^{-wx})^{-2}\cdot e^{-wx}\cdot x = \frac{1}{1+e^{-wx}}\cdot\frac{e^{-wx}}{1+e^{-wx}}\cdot x = \frac{1}{1+e^{-wx}}\cdot\left(1-\frac{1}{1+e^{-wx}}\right)\cdot x = \sigma(w)(1-\sigma(w))\cdot x$$

2.2 Error Function

† $E(w) = \frac{1}{2}\sum_{i=1}^{n}\left(\frac{1}{1+e^{-wx_i}} - y_i\right)^2 \to$

$$\frac{\partial E(w)}{\partial w} = \sum_{i=1}^{n}\left(\frac{1}{1+e^{-wx_i}} - y_i\right)\cdot\left(\frac{1}{1+e^{-wx_i}}\right)\cdot\left(1-\frac{1}{1+e^{-wx_i}}\right)\cdot x_i$$

3 Case : Log Loss

3.1 Log Loss

$$L(p) = C\cdot(-\ln p) + (1-C)\cdot(-\ln(1-p))$$

where $p(\mathbf{w}) = \frac{1}{1+e^{-\mathbf{wx}}}$.

3.2 Derivative of Log Loss

Recall that

$$\frac{\partial p}{\partial w_i} = p(1-p)x_i.$$

Now we calculate $\frac{\partial L}{\partial w_i}$ to update each w_i.

$$\frac{\partial L}{\partial w_i} = \frac{\partial L(p)}{\partial w_i} = \frac{\partial (C \cdot (-\ln p) + (1-C) \cdot (-\ln(1-p)))}{\partial w_i} = \frac{\partial (C \cdot (-\ln p) + (1-C) \cdot (-\ln(1-p)))}{\partial p} \cdot \frac{\partial p}{\partial w_i}$$

$$\begin{aligned}
&= \frac{\partial (C \cdot (-\ln p))}{\partial p} \cdot \frac{\partial p}{\partial w_i} + \frac{\partial ((1-C) \cdot (-\ln(1-p)))}{\partial p} \cdot \frac{\partial p}{\partial w_i} \\
&= C(-1)\frac{1}{p} \cdot \frac{\partial p}{\partial w_i} + (1-C)(-1)\frac{-1}{1-p} \cdot \frac{\partial p}{\partial w_i} \\
&= C(-1)\frac{1}{p} \cdot \frac{\partial p}{\partial w_i} + (1-C)\frac{1}{1-p} \cdot \frac{\partial p}{\partial w_i} \\
&= C(-1)\frac{1}{p} \cdot p(1-p)x_i + (1-C)\frac{1}{1-p} \cdot p(1-p)x_i \\
&= C(-1) \cdot (1-p)x_i + (1-C) \cdot px_i \\
&= -Cx_i + Cpx_i + px_i - Cpx_i \\
&= -Cx_i + px_i \\
&= (p-C)x_i.
\end{aligned}$$

$$\therefore \nabla L = (p-C)\mathbf{x}$$

3.3 Update

$$\mathbf{w_{n+1}} = \mathbf{w_n} - \eta \nabla L = \mathbf{w_n} - \eta(p-C)\mathbf{x} = \mathbf{w_n} - \eta\Big(\frac{1}{1+e^{-\mathbf{wx}}} - C\Big)\mathbf{x}.$$

Compare with the case of squared error function:

$$\mathbf{w_{n+1}} = \mathbf{w_n} - \eta(p-y)p(1-p)\mathbf{x} = \mathbf{w_n} - \eta\Big(\frac{1}{1+e^{-\mathbf{wx}}} - y\Big) \cdot \Big(\frac{1}{1+e^{-\mathbf{wx}}}\Big) \cdot \Big(1 - \frac{1}{1+e^{-\mathbf{wx}}}\Big) \cdot \mathbf{x}$$