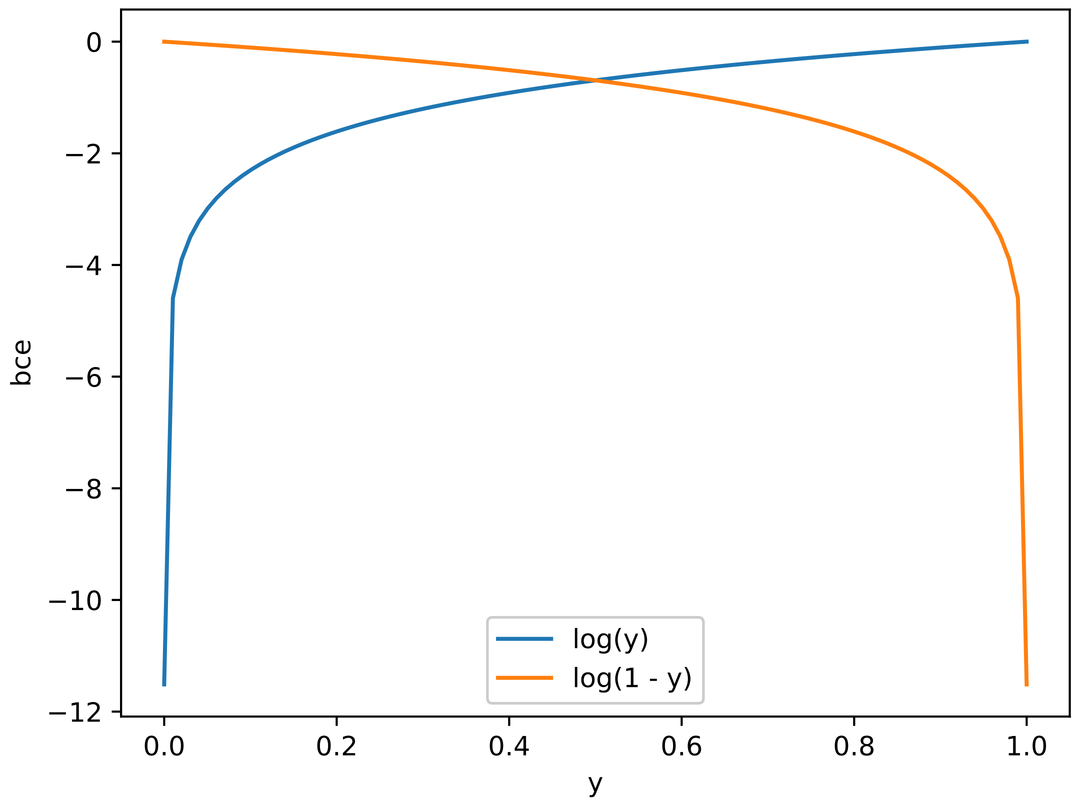
1 Theoretical part

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

1. $\sigma'(z) = \dfrac{e^{-z}}{(1 + e^{-z})^2} = \left(\frac{1}{\sigma(z)} - 1\right) \sigma^2(z) = (1 - \sigma(z))\sigma(z)$
2. $\sigma(-z) = \dfrac{1}{1 + e^{z}} = \dfrac{1}{e^{z}(e^{-z} + 1)} = \dfrac{e^{-z}}{1 + e^{-z}} = \dfrac{1 + e^{-z} - 1}{1 + e^{-z}} = 1 - \dfrac{1}{1 + e^{-z}} = 1 - \sigma(z)$
3. $h_w(x) = \sigma([1; x]^T w)$
4. BCE plot



5. Gradient

$$L(w) = -\frac{1}{N}\sum_{i=1}^{N}(y_{\{i\}}\log h_w(x_{\{i\}}) + (1-y_{\{i\}})\log(1-h_w(x_{\{i\}}))) + \alpha\sum_{j=1}^{N}(w_j)^2$$

$$\nabla_w L(w) = \left[\frac{\partial L(w)}{\partial w_0}, \frac{\partial L(w)}{\partial w_1}, \ldots, \frac{\partial L(w)}{\partial w_M}\right]$$

$$\text{For } k \neq 0:$$

$$\frac{\partial L(w)}{\partial w_k} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_{\{i\}}\frac{h'_w(x_{\{i\}})}{h_w(x_{\{i\}})}x_{\{i\},k} - (1-y_{\{i\}})\frac{h'_w(x_{\{i\}})}{1-h_w(x_{\{i\}})}x_{\{i\},k}\right] + 2\alpha w_k =$$

$$= -\frac{1}{N}\sum_{i=1}^{N}\left[y_{\{i\}}(1-h_w(x_{\{i\}}))x_{\{i\},k} - (1-y_{\{i\}})h_w(x_{\{i\}})x_{\{i\},k}\right] + 2\alpha w_k =$$

$$= -\frac{1}{N}\sum_{i=1}^{N}\left[y_{\{i\}}x_{\{i\},k} - y_{\{i\}}h_w(x_{\{i\}})x_{\{i\},k} - h_w(x_{\{i\}})x_{\{i\},k} + y_{\{i\}}h_w(x_{\{i\}})x_{\{i\},k}\right] + 2\alpha w_k =$$

$$= -\frac{1}{N}\sum_{i=1}^{N}\left[x_{\{i\},k}(y_{\{i\}} - h_w(x_{\{i\}}))\right] + 2\alpha w_k.$$

$$\frac{\partial L}{\partial w_0} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_{\{i\}} - h_w(x_{\{i\}})\right].$$

6. Gradient descent:

$$w_0(t+1) = w_0(t) + \frac{\lambda}{N}\sum_{i=1}^{N}(y_{\{i\}} - h_w(x_{\{i\}})),$$

$$w_k(t+1) = w_k(t) + \frac{\lambda}{N}\sum_{i=1}^{N}\left[x_{\{i\},k}(y_{\{i\}} - h_w(x_{\{i\}}))\right] - 2\alpha\lambda w_k(t).$$

7. Proof of only minimum.

$$bce(\hat{y}) = -y\log(\hat{y}) - (1-y)\log(1-\hat{y}),$$

$$bce'(\hat{y}) = -\frac{y}{\hat{y}} + \frac{1-y}{1-\hat{y}} = 0,$$

$$-y + y\hat{y} + \hat{y} - y\hat{y} = 0,$$

$$\hat{y} = y.$$

8. Minimization equivalence.

$$softplus(x) = \log(1 + e^x)$$

$$softplus(-tw^T x) = \log(1 + e^{-tw^T x}) = -\log\left(\frac{1}{1 + e^{-tw^T x}}\right) = \left\{z = w^T x\right\} - \log(\sigma(tz)),$$

Considering that $t \in \{-1, 1\}$:

$$-\log(\sigma(tz)) = -\frac{t+1}{2}\log(\sigma(z)) - \frac{1-t}{2}\log(\sigma(-z)).$$

According to item (2):

$$softplus(-tz) = -\frac{t+1}{2}\log(\sigma(z)) - \frac{1-t}{2}\log(1 - \sigma(z)).$$

After the change of variables ($t = 2y - 1$) we get:

$$softplus(-tz) = -y\log(\sigma(z)) - (1-y)\log(1 - \sigma(z)).$$