Explore the Sigmoid Function

In logistic regression, we assume that the conditional label distribution takes on the following form:

$$\mathrm{P}(y_i|\mathbf{x}_i;\mathbf{w}) = \frac{1}{1+e^{-y_i(\mathbf{w}^\top \mathbf{x}_i)}} = \sigma\left(y_i\left(\mathbf{w}^\top \mathbf{x}_i\right)\right)$$

where the sigmoid function is defined as:

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

You will explore the mathematical behavior of the sigmoid and discover properties that will prove useful in the optimization step. Answer the two questions that follow below. Once you think you've determined the solutions, click the buttons to reveal the answers.

1. Show that the sigmoid function $\sigma(\cdot)$ has the following property $\sigma(-z) = 1 - \sigma(z)$. By proving this property, you show that, for the binary classification problem, you have a properly defined a probabilistic model, namely:

$$\mathrm{P}(y_i = 1|\mathbf{x}_i) + \mathrm{P}(y_i = -1|\mathbf{x}_i) = 1$$

Click to reveal the answer

$$\begin{aligned}
\sigma(-z) &= \frac{1}{1+e^{z}} \\
&= \frac{e^{-z}}{e^{-z}(1+e^{z})} \\
&= \frac{e^{-z}}{e^{-z}+1} \\
&= \frac{e^{-z}+1-1}{e^{-z}+1} \\
&= \frac{e^{-z}+1}{e^{-z}+1} - \frac{1}{e^{-z}+1} \\
&= 1 - \frac{1}{e^{-z}+1} \\
&= 1 - \sigma(z)
\end{aligned}$$

2. Show the following property for its first derivative: $\sigma'(z) = \sigma(z)(1-\sigma(z))$.

Click to reveal the answer

$$
\begin{aligned}
\sigma'(z) &= \frac{d}{dz}\left(\frac{1}{1+e^{-z}}\right) \\
&= \frac{-1}{(1+e^{-z})^2} \cdot \frac{d}{ds}(1+e^{-z}) \\
&= \frac{-1}{(1+e^{-z})^2} \cdot (-e^{-z}) \\
&= \frac{e^{-z}}{(1+e^{-z})^2} \\
&= \frac{1}{1+e^{-z}} \cdot \frac{e^{-z}}{1+e^{-z}} \\
&= \frac{1}{1+e^{-z}} \cdot \frac{e^{-z}+1-1}{1+e^{-z}} \\
&= \frac{1}{1+e^{-z}} \cdot \left(\frac{e^{-z}+1}{1+e^{-z}} - \frac{1}{1+e^{-z}}\right) \\
&= \frac{1}{1+e^{-z}} \cdot \left(1 - \frac{1}{1+e^{-z}}\right) \\
&= \sigma(z)(1-\sigma(z))
\end{aligned}
$$