Variational formulation of the sigmoid function

Caio Corro

In this note, I derive the variational formulation of the sigmoid function used in *Computing upper and lower bounds on likelihoods in intractable networks* (Jaakkola and Jordan).

The sigmoid function and its derivative are defined as follows:

$$\sigma(x) = \frac{\exp(x)}{1+\exp(x)} = \frac{1}{1+\exp(-x)}$$
$$\sigma'(x) = \sigma(x)(1-x)$$

This function is neither convex or concave. However, the function

$$f(x) = -\log \sigma(x) = -\log \frac{\exp(x)}{1+\exp(x)} = -x + \log(1+\exp(x))$$

is convex. We can check this by computing the second order derivative:

$$\frac{\partial}{\partial^2 x} - \log \sigma(x) = -\frac{\partial}{\partial x}\frac{1}{\sigma(x)}\sigma(x)(1-x) = -\frac{\partial}{\partial x}(1-x) = 1$$

which is positive, so f is convex. Therefore, we can build a variational approximation of f using its Fenchel biconjugate:

$$f(x) = f^{**}(x) = \max_y xy - f^*(y)$$

We first compute the conjugate of f:

$$f^*(y) = \max_x xy - f(x) = \max_x xy + \log \sigma(x)$$

Using first order optimality condition, we have:

$$\frac{\partial}{\partial x}(xy + \log \sigma(x)) = 0$$
$$y + \frac{1}{\sigma(x)}\sigma(x)(1-\sigma(x)) = 0$$
$$y = \sigma(x) - 1$$

meaning that $\text{dom}\, f^* = (-1, 0)$. The inverse of sigmoid function is defined as $\sigma^{-1}(x) = \log \frac{x}{1-x}$, therefore:

$$x = \log \frac{y+1}{-y}$$

Hence, the Fenchel conjugate of f is:

$$\begin{aligned}
f^*(y) &= y \log \frac{y+1}{-y} + \log \sigma\left(\log \frac{y+1}{-y}\right) \\
&= y\log(y+1) - y\log(-y) + \log\frac{y+1}{-y} - \log\left(1 + \frac{y+1}{-y}\right) \\
&= y\log(y+1) - y\log(-y) + \log(y+1) - \log(-y) - \log\left(\frac{-y}{-y} + \frac{y+1}{-y}\right) \\
&= (y+1)\log(y+1) - y\log(-y) - \log(-y) - \log\left(\frac{1}{-y}\right) \\
&= (y+1)\log(y+1) - y\log(-y) - \log(-y) - \log 1 + \log(-y) \\
&= (y+1)\log(y+1) - y\log(-y)
\end{aligned}$$

We can check that biconjugacy holds:

$$f^{**}(x) = \max_y xy - f^*(y) = \max_y xy - (y+1)\log(y+1) + y\log(-y)$$

By first order optimality conditions:

$$\begin{aligned}
\frac{\partial}{\partial y}(xy - (y+1)\log(y+1) + y\log(-y)) &= 0 \\
x - \log(y+1) - 1 + \log(-y) + 1 &= 0 \\
\log\frac{y+1}{-y} &= x \\
\log\frac{y+1}{1-(y+1)} &= x \\
y+1 &= \sigma(x) \\
y &= \sigma(x) - 1
\end{aligned}$$

Hence:

$$\begin{aligned}
f^{**}(x) &= x(\sigma(x)-1) - (\sigma(x)-1+1)\log(\sigma(x)-1+1) + (\sigma(x)-1)\log(-\sigma(x)+1) \\
&= x\sigma(x) - x - \sigma(x)\log\sigma(x) + (\sigma(x)-1)\log(-\frac{\exp(x)}{1+\exp(x)} + \frac{1+\exp(x)}{1+\exp(x)}) \\
&= x\sigma(x) - x - \sigma(x)\log\sigma(x) + (\sigma(x)-1)\log(\sigma(-x)) \\
&= x\sigma(x) - x - \sigma(x)\log\sigma(x) + (\sigma(x)-1)\log(\frac{1}{1+\exp(x)}) \\
&= x\sigma(x) - x - \sigma(x)\log\sigma(x) + (\sigma(x)-1)(\log 1 - \log(1+\exp(x))) \\
&= x\sigma(x) \underbrace{-x} - \sigma(x)\log\sigma(x) - \sigma(x)\log(1+\exp(x))) + \underbrace{\log(1+\exp(x)))} \\
&= \underbrace{-\log\sigma(x)} + x\sigma(x) - \sigma(x)\log\sigma(x) - \sigma(x)\log(1+\exp(x))) \\
&= -\log\sigma(x) + x\sigma(x) - \sigma(x)\log\sigma(x) - \sigma(x)\left(x \underbrace{-x + \log(1+\exp(x))}_{-\log\sigma(x)}\right) \\
&= -\log\sigma(x) \\
&= f(x)
\end{aligned}$$

We can now the variational formulation of the sigmoid function:

$$\begin{aligned}
\sigma(x) &= \exp(-f(x)) \\
&= \exp(-f^{**}(x)) \\
&= \exp(-\max_{y\in(-1,0)} xy - (y+1)\log(y+1) + y\log(-y)) \\
&= \exp(\min_{y\in(-1,0)} -xy + (y+1)\log(y+1) - y\log(-y)) \\
&= \exp(\min_{y\in(0,1)} xy + (1-y)\log(1-y) + y\log(y)) \\
&= \exp(\min_{y\in(0,1)} xy - H(y)) \\
&= \min_{y\in(0,1)} \exp(xy - H(y))
\end{aligned}$$

where H is the Fermi-Diract entropy function (i.e. entropy of a Bernoulli distribution).