CS/CS 316/365 Deep Learning

Activity 4

September 23, 2024

Loss Functions

Activity needs to be handwritten. Submission will be online on canvas only.

- Show that the logistic sigmoid function sig[z] becomes 0 as z $\rightarrow$ -∞, is 0.5 when z = 0, and becomes 1 when z $\rightarrow$ ∞. Sigmoid's equation is given below. Showing doesn't require lengthy proof. Try putting in these values and show that result reaches to where it should be.

$$\text{sig}[z] = \frac{1}{1 + \exp[-z]}$$

  Solution:
  When z = ∞, the exponential term become infinitely large and so the fraction is zero. When z = +∞, the exponential term becomes zero and so the fraction is one. When z = 0, the exponential term becomes one, and so the fraction is 1/(1 + 1) = 0.5.

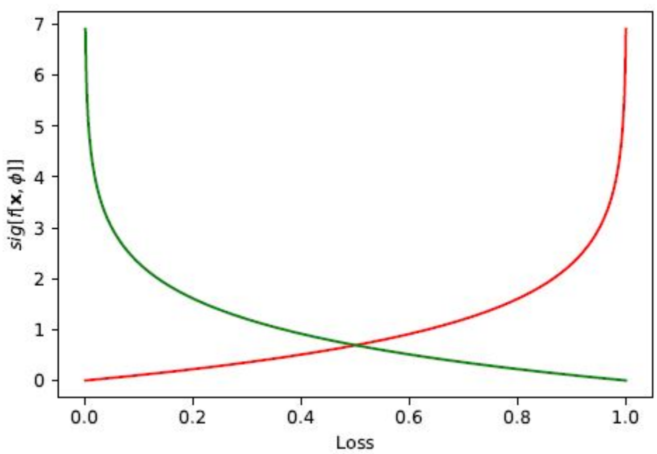
- The loss L for binary classification for a single training pair {x, y} is:

$$L = -(1-y)\log\Big[1 - \text{sig}[\text{f}[\mathbf{x}, \boldsymbol{\phi}]]\Big] - y\log\Big[\text{sig}[\text{f}[\mathbf{x}, \boldsymbol{\phi}]]\Big]$$

  where sig[z] is given above in first task. Plot this loss as a function of the transformed network output sig[f[x,ϕ]] $\in$ [0, 1] when the training label

  - y = 0
  - y = 1

  Solution:
  The loss as a function of transformed network output. Red curve is for case where y = 0 and green curve is for y = 1.



- Consider a multivariate regression problem where we predict ten outputs, so y $\in \mathbb{R}^{10}$, and model each with an independent normal distribution where the means μ_d are predicted by the network, and variances σ^2 are constant. Write an expression for the likelihood Pr(y | f[x,ϕ]). Show that minimizing the negative log-likelihood of this model is still equivalent to minimizing a sum of squared terms if we don't estimate the variance σ^2.

  Solution:
  The likelihood is:

$$Pr(y|\mathrm{f}[\mathbf{x},\phi],\sigma^2) = \prod_{d=1}^{10} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_d - \mathrm{f}_d[\mathbf{x},\phi])^2}{2\sigma^2}\right]$$

  The loss function is:

$$L = \sum_{i=1}^{I} \log\left[\prod_{d=1}^{10} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_{id} - \mathrm{f}_d[\mathbf{x}_i,\phi])^2}{2\sigma^2}\right]\right]$$

  This can be simplified to:

$$\begin{aligned}
\hat{\phi} &= \underset{\phi}{\operatorname{argmin}}\left[-\sum_{i=1}^{I}\sum_{d=1}^{10} \log\left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_{id} - \mathrm{f}_d[\mathbf{x}_i,\phi])^2}{2\sigma^2}\right]\right]\right] \\
&= \underset{\phi}{\operatorname{argmin}}\left[-\sum_{i=1}^{I}\sum_{d=1}^{10} \log\left[\frac{1}{\sqrt{2\pi\sigma^2}}\right] - \frac{(y_{id} - \mathrm{f}_d[\mathbf{x}_i,\phi])^2}{2\sigma^2}\right] \\
&= \underset{\phi}{\operatorname{argmin}}\left[-\sum_{i=1}^{I}\sum_{d=1}^{10} -\frac{(y_{id} - \mathrm{f}_d[\mathbf{x}_i,\phi])^2}{2\sigma^2}\right] \\
&= \underset{\phi}{\operatorname{argmin}}\left[\sum_{i=1}^{I}\sum_{d=1}^{10} (y_{id} - \mathrm{f}_d[\mathbf{x}_i,\phi])^2\right],
\end{aligned}$$