Knowledge Check 5 – Loss Functions

DS598 B1 – DL4DS

Spring, 2024

Problem 5.2 The loss L for binary classification for a single training pair x, y is:

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

where $sig[\bullet]$ is defined in equation 5.18. Plot this loss as a function of the transformed network output $sig[f[x,\phi]] \in [0,1]$ (i) when the training label $y = 0$ and (ii) when $y = 1$.

Problem 5.3 *Optional problem, but recommend doing it. Hint: you should be able to answer in 3-4 lines of math. You don't have define what the Bessel function is. Just leave it as a symbol.*

Problem 5.6 Consider building a model to predict the number of pedestrians $y \in 0, 1, 2, ...$ that will pass a given point in the city in the next minute, based on data x that contains information about the time of day, the longitude and latitude, and the type of neighborhood. A suitable distribution for modeling counts is the Poisson distribution (figure 5.15 from book). This has a single parameter $\lambda > 0$ called the rate that represents the mean of the distribution. The distribution has probability density function:

$$Pr(y = k) = \frac{\lambda^k e^{-\lambda}}{k!}.$$

Use the recipe in section 5.2 to design a loss function for this model assuming that we have access to I training pairs $\{xi, yi\}$.

Problem 5.7 Consider a multivariate regression problem where we predict 10 outputs so $\mathbf{y} \in R^{10}$ and model each with an independent normal distribution where the means μ_d are predicted by the network, and variances σ^2 are all the same. Write an expression for the likelihood $Pr(\mathbf{y}|\mathbf{f}[\mathbf{x}, \phi])$ for this model. 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.