Homework 3: Loss Functions Worksheet

DS542

Spring 2026

In this worksheet, you will derive some of the common loss functions for regression and classification problems. Feel free to write your answers by hand. No need for LaTeX; just make sure it is legible.

1. **From Joint Probability to Negative Log-Likelihood Loss Function**

   In the lecture on loss and Chapter 5 in UDL, we stated that we are searching for a probability distribution that maximizes the probability of the target values given the input values, or

   $$Pr(y_1, y_2, \ldots, y_I | x_1, x_2, \ldots, x_I).$$

   From there we ended up with an estimate for the parameters of a model that minimze the negative log-likelihood loss function:

   $$\hat{\phi} = \arg\min_{\phi} \left[-\sum_{i=1}^{I} \log\left[Pr(\mathbf{y}_i | \mathbf{f}[\mathbf{x}_i, \phi]) \right] \right].$$

   (5 points) Recreate all the steps to go from the joint probability to the negative log-likelihood loss function.

2. **Negative Log-Likelihood Loss Function for Gaussian Distribution**

If we are performing univariate regression and assume a gaussian probability distribution where the standard deviation, σ, is fixed, the negative log-likelihood loss function is given by:

$$
\begin{aligned}
L[\phi] &= -\sum_{i=1}^{I} \log \left[Pr\left(y_i | f[x_i, \phi], \sigma^2\right)\right] \\
&= -\sum_{i=1}^{I} \log \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_i - f[x_i, \phi])^2}{2\sigma^2}\right]\right]
\end{aligned}
$$

(5 points) Show step by step that the equation above simplifies to the sum of squared errors loss function.

3. For binary classification we can use the Bernoulli distribution to model the probability of the target values:

$$Pr(y|\lambda) = (1 - \lambda)^{1-y} \cdot \lambda^y$$

As stated in the lecture and the textbook, we can train a model to predict the parameter λ and then use it to predict the probability of the target values.

$$Pr(y|\mathbf{x}) = (1 - \text{sig}[f[\mathbf{x}|\phi]])^{1-y} \cdot \text{sig}[f[\mathbf{x}|\phi]]^y$$

$$\hat{\lambda} = \arg \max_{\lambda} \left[\prod_{i=1}^{I} Pr(y_i|\lambda) \right]$$

(5 points) Derive the negative log-likelihood loss function for the Bernoulli distribution.

4. **(5 points)** For multiclass classification, define the negative log-likelihood loss function in terms of the softmax function and the model $\mathbf{f}[\mathbf{x}, \phi]$ and then how the loss function can be simplified when we apply the log to the softmax equation.