Assignment: Understanding Loss Functions in Supervised Learning

Due Date: [Insert Due Date Here]

Introduction:

In supervised learning, the choice of a loss function is a critical step that dictates how our model learns and what objective it optimizes. Building upon our understanding of supervised learning as (1) choosing a representation, (2) choosing a loss function, and (3) minimizing the loss, this assignment will delve deeper into the concept and application of various loss functions, particularly in the context of classification.

Instructions: Please answer the following questions thoroughly, referencing the provided lecture notes where appropriate.

Part 1: Fundamental Concepts of Loss Functions

Defining Loss Functions:

What is the primary purpose of a loss function in supervised learning?

How does a loss function relate to the overall goal of minimizing empirical risk?

Loss Functions Across Problem Types:

The notes discuss different types of supervised learning problems. For each of the following, identify the typical target space (Y) and provide the formula for a commonly used loss function mentioned in the text:

Linear Regression

Binary Classification (using $y \in \{-1, 1\}$)

Multiclass Classification

Part 2: Margin-Based Loss Functions for Binary Classification

The Concept of Margin:

In binary classification (with labels $y \in \{-1, 1\}$), what is the "margin" for an example (x, y) with respect to a parameter vector θ? Provide its mathematical expression.

Explain why a large positive margin is desirable for a correctly classified example.

Desired Characteristics of Classification Loss:

Describe the general shape and behavior (as depicted in Figure 1 of the notes) that we desire for a classification loss function $\varphi(z)$ with respect to the margin z. Specifically, how should $\varphi(z)$ behave as $z \to \infty$ and as $z \to -\infty$?

The Zero-One Loss:

Provide the formula for the zero-one loss, $\varphi zo(z)$.

From an intuitive perspective, why might the zero-one loss seem like the "most natural" choice for classification?

Despite its intuitive appeal, the notes state that φzo is "discontinuous, non-convex, and NP-hard to minimize." Briefly explain why these characteristics make it problematic for practical optimization in machine learning.

Part 3: Common Classification Loss Functions

Comparing Practical Loss Functions:

The notes introduce three commonly used margin-based loss functions that overcome the limitations of the zero-one loss. For each of the following, provide its mathematical formula and briefly describe its key characteristics or behavior with respect to the margin ($z = yxT\theta$), referencing Figure 2:

(i) Logistic Loss (φlogistic(z))

(ii) Hinge Loss (φhinge(z))

(iii) Exponential Loss (φexp(z))

Loss Functions and Algorithms:

The choice of a loss function often gives rise to specific machine learning algorithms. Which well-known machine learning algorithms are associated with:

The hinge loss (φ_{hinge})?

The exponential loss (φ_{exp})?

Part 4: Empirical Risk Minimization

Empirical Risk for Margin-Based Losses:

Given a training set of pairs $\{(x^{(i)}, y^{(i)})\}_{i=1}^{m}$ and a margin-based loss function $\varphi(z)$, write down the general formula for the empirical risk $J(\theta)$ that we aim to minimize.

Explain how minimizing this empirical risk helps us find an optimal parameter vector θ for our model.