Problem Set 5 – Loss Functions

DS 542 – DL4DS

Fall, 2024

The following problems assume a data set with one input x and one output y. The input x is a binary variable only taking on values zero and one, and $Pr(y|x = i)$ is normally distributed with mean μ_i and standard deviation σ_i. In the training data, $Pr(x = 1) = p$. Both the μ_i's and σ_i's are unknown.

Problem 1 If we model this distribution as $y = \phi_0 + \phi_1 x$ and choose the L_2 loss function, what is the loss function $L[]$ as parameterized by ϕ_0, ϕ_1, and p?

Problem 2

What are the gradients of the loss loss function $L[]$ with respect to ϕ_0 and ϕ_1? Assume the data set is large enough that the analytical solution is accurate (i.e. ignore statistical variations not explicitly modeled).

Problem 3

If you train this model with gradient descent to the global minimum loss, what will the final values of parameters ϕ_0 and ϕ_1 be? Write your answer in terms of the μ_i's and σ_i's and p. Again, assume the data set is large enough that the analytical solution is accurate.

Problem 4

Are there any values of ϕ_0 and ϕ_1 where gradient descent could get stuck? That is, are there any parameter choices where the gradients become zero, but they are not at the global minimum?