Problem Set 5 - Regularization11111

DS542 – DL4DS

Spring, 2025

Note: Refer to the equations in the *Understanding Deep Learning* textbook to solve the following problems.

AI/Correction Statement (1 point)

You may use ChatGPT/Generative AI as a resource to help you complete the assignment. However, it must be used constructively to help you understand things you are unsure of, and be built upon with original work.

You must cite your interaction by describing your prompt and the corresponding response. In addition, you must explain all output from the AI that you implement in your assignment. Failure to do so could result in credit deduction.

The official GAIA Policy can be found here.

Moreover, if this is a correction submission after the initial submission, you must provide a reflection on what you learned from the initial submission and how you corrected it.

Problem 9.1 (4 points)

Consider a model where the prior distribution over the parameters is a normal distribution with mean zero and variance σ_ϕ^2 so that

$$Pr(\phi) = \prod_{j=1}^{J} \mathcal{N}(\phi_j | 0, \sigma_\phi^2), \tag{1}$$

where j indexes the model parameters. When we apply a prior, we maximize

$$\prod_{i=1}^{I} Pr(y_i | x_i, \phi) Pr(\phi). \tag{2}$$

Show that the associated loss function of this model is equivalent to L2 regularization.

Problem 9.5 (4 points)

Show that the weight decay parameter update with decay rate λ:

$$\phi \leftarrow (1 - \lambda)\phi - \alpha \frac{\partial L}{\partial \phi}, \tag{3}$$

on the original loss function $L[\phi]$ is equivalent to a standard gradient update using L2 regularization, so that the modified loss function $\tilde{L}[\phi]$ is:

$$\tilde{L}[\phi] = L[\phi] + \frac{\lambda}{2\alpha} \sum_k \phi_k^2, \tag{4}$$

where ϕ represents the parameters, and α is the learning rate.