Problem Set 3 – Loss Functions and Fitting Models

DS542 – DL4DS

Spring, 2025

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

Problem 5.9

Consider a multivariate regression problem in which we predict the height of an individual in meters and their weight in kilos from some data x. Here, the units take quite different values. What problems do you see this causing? Propose two solutions to these problems.

Problem 6.6

Which of the functions in Figure 6.11 from the book is convex? Justify your answer. Characterize each of the points 1–7 as (i) a local minimum, (ii) the global minimum, or (iii) neither.

Problem 6.10

Show that the momentum term m_t (equation (6.11)) is an infinite weighted sum of the gradients at the previous iterations and derive an expression for the coefficients (weights) of that sum.

Problem 5.9

Consider a multivariate regression problem in which we predict the height of an individual in meters and their weight in kilos from some data x. Here, the units take quite different values. What problems do you see this causing? Propose two solutions to these problems.

Having variables with different scales

Problems: ① weight of the model imbalanced.

② loss function highly skewed.

↳ slowing down the gradient descent make convergence more difficult.

Solutions: ① Normalized scale (Min-Max)

② standardization (z-score normalization)

Problem 6.6

Which of the functions in Figure 6.11 from the book is convex? Justify your answer. Characterize each of the points 1–7 as (i) a local minimum, (ii) the global minimum, or (iii) neither.

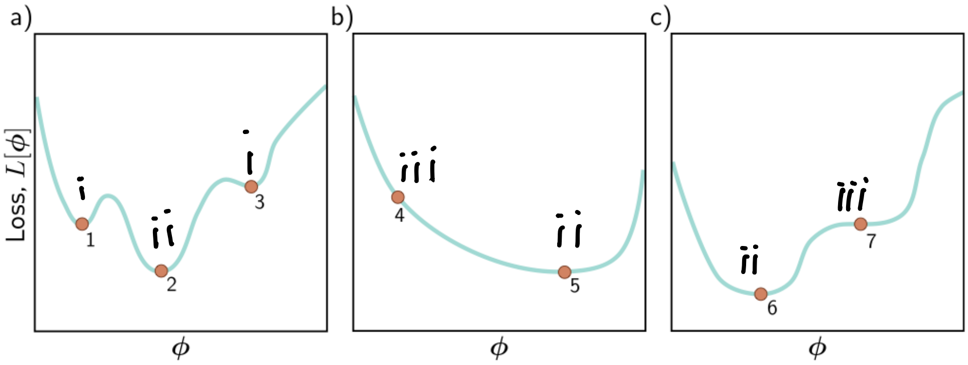

Figure 6.11 Three 1D loss functions for problem 6.6.

b) is convex

Problem 6.10

Show that the momentum term m_t (equation (6.11)) is an infinite weighted sum of the gradients at the previous iterations and derive an expression for the coefficients (weights) of that sum.

$$\nabla_t = \sum_{i \in \mathcal{B}_t} \frac{\partial \ell_i[\phi_t]}{\partial \phi}$$

$$\begin{aligned}
\mathbf{m}_{t+1} &\leftarrow \beta \cdot \mathbf{m}_t + (1-\beta) \sum_{i \in \mathcal{B}_t} \frac{\partial \ell_i[\phi_t]}{\partial \phi} \\
\phi_{t+1} &\leftarrow \phi_t - \alpha \cdot \mathbf{m}_{t+1},
\end{aligned}$$

$m_{t+1} = \beta m_t + (1-\beta)\nabla_t$

$\beta(\beta m_{t-1} + (1-\beta)\nabla_{t-1})(1-\beta)\nabla_t$

$\beta(\beta(\beta m_{t-2} + (1-\beta)\nabla_{t-2})(1-\beta)\nabla_{t-1})(1-\beta)\nabla_t$

$\beta(\beta(\beta(\beta m_{t-3} + (1-\beta)\nabla_{t-3})(1-\beta)\nabla_{t-2})(1-\beta)\nabla_{t-1})(1-\beta)\nabla_t$

$\vdots$

$\beta^t m_0 (1-\beta)\sum_{i=0}^{t-1} \beta^i \nabla_{t-i}$