Python and Maths Basics

Exercise 1 : Converting between model function space and the loss landscape

Consider the loss $l_2(c, y(\mathbf{x}))$ as defined in the homework sheet for the point $(4, 6)$.

(a) By looking at the model function space (the 2D coordinate system in which the model function is represented as a straight line), intuitively explain why $l_2 = 0$ holds for a set of more than one instance of model parameters.

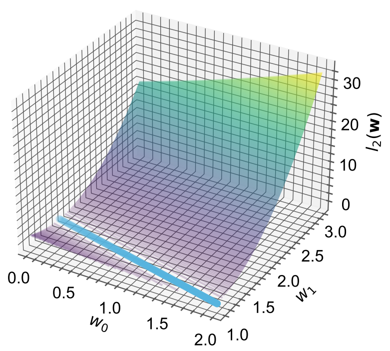
Answer

The loss l_2 is optimal (i.e. 0) for multiple possible parameter configurations (i.e. straight lines in the function space that go through the point).

(b) Show through calculations that $l_2 = 0$ is a straight line in the loss landscape.

Answer

$$\begin{aligned} l_2(c, y(\mathbf{x})) &= 0 \\ \frac{1}{2} \cdot (6 - w_0 - 4w_1)^2 &= 0 \\ 6 - w_0 - 4w_1 &= 0 \\ w_1 &= -\frac{1}{4}w_0 + \frac{3}{2} \end{aligned}$$



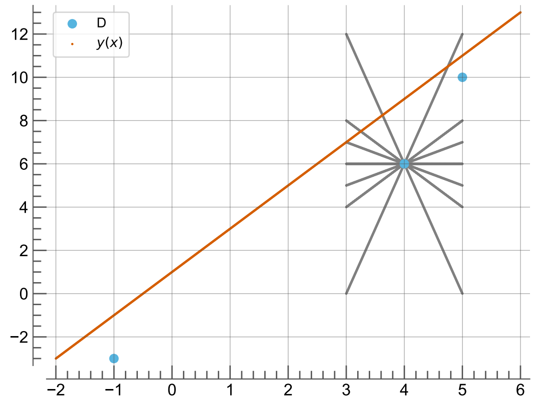
(c) To what structure does this correspond in the function space? Plot and try to show your conjecture using calculations. Hint: Compute $y(4)$ and $\frac{\partial y}{\partial x}$.

Answer

$$\begin{aligned} y(x) &= w_0 + w_1 x \\ &= w_0 + (-\frac{1}{4}w_0 + \frac{3}{2})x \\ y(4) &= w_0 + (-w_0 + 6) \\ &= 6 \\ &= c \\ \frac{\partial y}{\partial x} &= -\frac{1}{4}w_0 + \frac{3}{2} \end{aligned}$$

This is const. w.r.t. x, and has $]-\infty, +\infty[$ as its range of values w.r.t. w_0.

Thus, we get all straight lines through $(4, 6)$.



Exercise 2 : Advanced model functions

What if we want to fit a parabola instead of a straight line?

(a) Define the model function.

Answer

$$y(x) = w_2 x^2 + w_1 x + w_0$$

(b) How can we find $(\mathbf{x}, c)$, given the points to fit through (e.g., $(4, 6), (-1, -3), (5, 10)$)? Hint: Write the model function as a vector dot product $\mathbf{w}^T \cdot \mathbf{x}$.

Answer

$$y(\mathbf{x}) = \mathbf{w}^T \cdot \mathbf{x}$$
$$\mathbf{w} = \begin{pmatrix} w_0 \\ w_1 \\ w_2 \end{pmatrix}$$
$$\mathbf{x} = \begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix}$$

c is again simply the second coordinate of each point.

(c) Vaguely describe the loss landscape.

Answer

The loss (landscape) has three parameters (w_0, w_1, w_2). Given two of those parameters, there exists a value for the third parameter so that $l_2 = 0$.

Exercise 3 : Gradient descent and loss functions

Consider the general case, but you might want to check back on the loss landscape plot in the homework exercise.

(a) In which direction does the gradient point?

Answer

The gradient points in the direction of greatest increase (or steepest ascent).

(b) In which direction does the negative gradient point?

Answer

The negative gradient points in the direction of greatest decrease (or steepest descent).

(c) Why does that help in the context of a loss landscape?

Answer

We optimize (i.e. minimize) the loss on the most direct way.

Exercise 4 : Limits of LMS

(a) What happens to the loss landscape in further iterations of the LMS algorithm?

Answer

The loss landscape gets completely redefined due to it being defined for a single point in each iteration only.

(b) Why is that a problem?

Answer

This might lead to slow or no convergence due to oscillation between (partly) conflicting values.

(c) What could be the solution?

Answer

Defining a global loss (will soon be done in the lecture).