Mathematics of Big Data & Machine Learning

Hayden Jananthan & Jeremy Kepner

18.095, January 10, 2024

1. Recall that a ring $\mathsf{R} = (R, +, \cdot, -, 0, 1)$ consists of a set R equipped with two binary operations $+$ and $\cdot$, a unary operation $-$, and two constant 0 and 1 satisfying the equational laws:

$$\begin{aligned} x + (y + z) &= (x + y) + z, & x \cdot (y \cdot z) &= (x \cdot y) \cdot z, \\ x + y &= y + x, & x \cdot 1 = 1 \cdot x &= x, \\ x + 0 = 0 + x &= 0, & x \cdot (y + z) &= (x \cdot y) + (x \cdot z), \\ x + (-x) = (-x) + x &= 0, & (y + z) \cdot x &= (y \cdot x) + (z \cdot x). \end{aligned}$$

   Show that in any ring $\mathsf{R} = (R, +, \cdot, -, 0, 1)$, 0 is a *multiplicative annihilator*, i.e., $0 \cdot x = x \cdot 0 = 0$ for all $x \in R$.

2. Consider a feed-forward neural network with n input neurons and m output neurons and no hidden layers. The inference step/forward propagation for this network can be written as

$$I(\mathbf{x}) = h(\mathbf{W}\mathbf{x} + \mathbf{b})$$

   where $\mathbf{W} = (w_{i,j})_{i,j}$ is the $m \times n$ weight matrix, $\mathbf{b} = (b_1, b_2, \ldots, b_m) \in \mathbb{R}^m$ is the bias vector, $\mathbf{x} \in \mathbb{R}^n$ is our input vector, and $h : \mathbb{R} \to \mathbb{R}$ is our activation function (applied elementwise when applied to a vector). You may assume that h is differentiable.

   Let X denote the set of training examples and let $y(\mathbf{x})$ be the ground-truth output associated with $\mathbf{x} \in X$. The cost function associated with this network is given by

$$C = \frac{1}{2n} \sum_{\mathbf{x} \in X} \| I(\mathbf{x}) - y(\mathbf{x}) \|^2,$$

   where n is the size of X. Find $\frac{\partial C}{\partial w_{i,j}}$ and $\frac{\partial C}{\partial b_i}$ for each $1 \le i \le m$ and $1 \le j \le n$.

3. Define

$$\mathbf{A} := \begin{array}{c} \\ a \\ b \\ c \end{array} \begin{array}{c} \begin{array}{ccc} a & b & d \end{array} \\ \begin{bmatrix} 3 & 2 & -1 \\ -1 & 1 & \\ & 2 & -1 \end{bmatrix} \end{array} \qquad \mathbf{B} := \begin{array}{c} \\ a \\ c \\ d \end{array} \begin{array}{c} \begin{array}{ccc} a & b & e \end{array} \\ \begin{bmatrix} -1 & 1 & 2 \\ 3 & -3 & \\ 1 & 3 & -1 \end{bmatrix} \end{array}$$

   Calculate $\mathbf{A} \oplus \mathbf{B}$, $\mathbf{A} \otimes \mathbf{B}$, and $\mathbf{A} \oplus.\!\otimes \mathbf{B}$ when $\oplus$ and $\otimes$ are given by:

   (a) $\oplus = +$, $\otimes = \cdot$

   (b) $\oplus = \min$, $\otimes = \max$

   (c) $\oplus = \max$, $\otimes = +$

   (Note: Be sure to interpret empty entries based on the choice of $\oplus$.)