ENGR 3321: Introduction to Deep Learning for Robotics

Neural Network 201: Two-Input One-Output Function

09/25/2023

Outline

- Forward Pass
- Loss Function
- Gradient Descent
- Linear Model

Two-Input One-Output Linear Function

Assume a dataset contains M objects

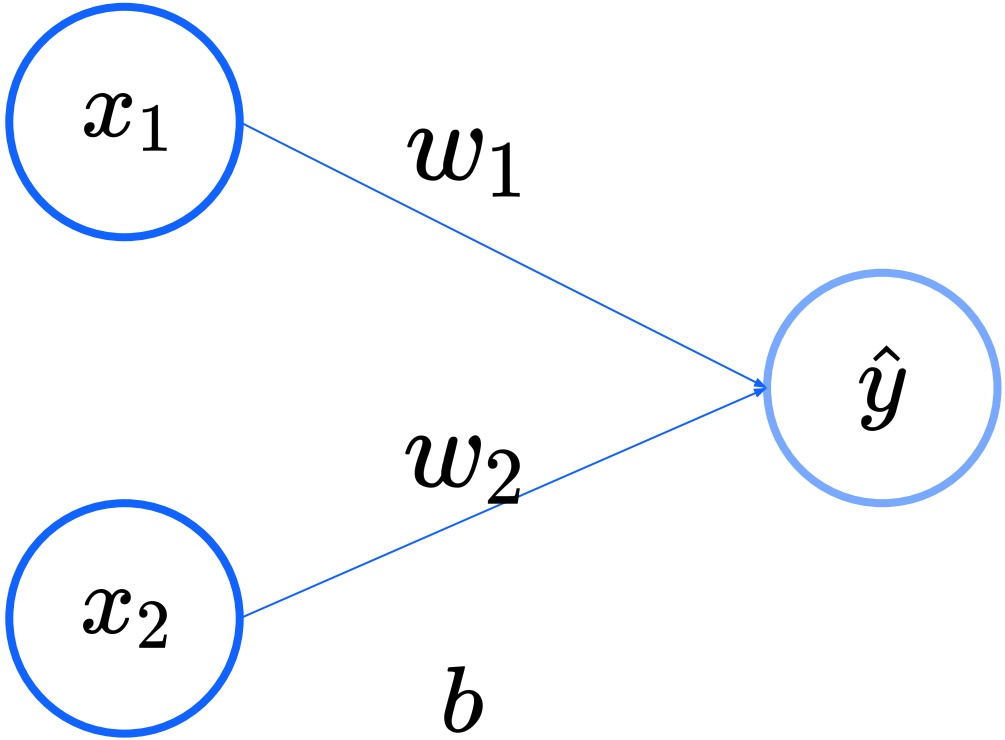
Each object can be described by two features: x_1 and x_2.

Each object has a property: y

We would like to know the relationship between x_1, x_2 and y

$$f(x_1, x_2) = \hat{y} = w_1 x_1 + w_2 x_2 + b$$

Neural Network Form



Matrix Form

$$\hat{\mathbf{y}} = \mathbf{X} \cdot \mathbf{w}^{\mathbf{T}} + \mathbf{b}$$

Input (Feature) Matrix

$$\mathbf{x_1} = \begin{bmatrix} {}^{(1)}x_1 \\ {}^{(2)}x_1 \\ \cdot \\ \cdot \\ \cdot \\ {}^{(M)}x_1 \end{bmatrix}_{(M,1)} \quad \mathbf{x_2} = \begin{bmatrix} {}^{(1)}x_2 \\ {}^{(2)}x_2 \\ \cdot \\ \cdot \\ \cdot \\ {}^{(M)}x_2 \end{bmatrix}_{(M,1)} \quad \mathbf{X} = \begin{bmatrix} \mathbf{x_1} & \mathbf{x_2} \end{bmatrix}_{(M,2)}$$

Model Parameters

$$\mathbf{w} = \begin{bmatrix} w_1 & w_2 \end{bmatrix}_{(1,2)} \qquad \mathbf{b} = \begin{bmatrix} b \\ b \\ . \\ . \\ . \\ b \end{bmatrix}_{(M,1)}$$

Mean Square Error Loss

$$\mathcal{L}(\hat{\mathbf{y}}, \mathbf{y}) = \frac{1}{M} \sum_{i=1}^{M} \left({}^{(i)}\hat{y} - {}^{(i)} y \right)^2 = \frac{1}{2} \Sigma (\hat{\mathbf{y}} - \mathbf{y})^2$$

Gradient of Loss

$$\nabla \mathcal{L} = \begin{bmatrix} \frac{\partial \mathcal{L}}{\partial w_1} & \frac{\partial \mathcal{L}}{\partial w_2} & \frac{\partial \mathcal{L}}{\partial b} \end{bmatrix}$$

$$\frac{\partial \mathcal{L}}{\partial w_1} = \frac{\partial \mathcal{L}}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = \frac{1}{M} \sum_{i=1}^{M} \left({}^{(i)}\hat{y} - {}^{(i)}y \right) .^{(i)} x_1$$

$$\frac{\partial \mathcal{L}}{\partial w_2} = \frac{\partial \mathcal{L}}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2} = \frac{1}{M} \sum_{i=1}^{M} \left({}^{(i)}\hat{y} - {}^{(i)}y \right) .^{(i)} x_2$$

$$\frac{\partial \mathcal{L}}{\partial b} = \frac{\partial \mathcal{L}}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b} = \frac{1}{M} \sum_{i=1}^{M} \left({}^{(i)}\hat{y} - {}^{(i)}y \right)$$

Vectorized Gradient

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \begin{bmatrix} \frac{\partial \mathcal{L}}{\partial w_1} & \frac{\partial \mathcal{L}}{\partial w_2} \end{bmatrix} = \frac{1}{M}(\hat{\mathbf{y}} - \mathbf{y})^T \cdot \mathbf{X}$$

$$\frac{\partial \mathcal{L}}{\partial b} = \frac{1}{M}\Sigma(\hat{\mathbf{y}} - \mathbf{y})$$

Vectorized Gradient Descent

Given dataset: $\left\{ \left({}^{(1)}\mathbf{x}, {}^{(1)}y \right), \left({}^{(2)}\mathbf{x}, {}^{(2)}y \right), \ldots, \left({}^{(M)}\mathbf{x}, {}^{(M)}y \right) \right\}$

Initialize $\mathbf{w}$ *and* b

Repeat until converge {

$$\mathbf{w} := \mathbf{w} - \alpha \frac{\partial \mathcal{L}}{\partial \mathbf{w}}$$

$$b := b - \alpha \frac{\partial \mathcal{L}}{\partial b}$$

}

where α is learning rate