ENGR 3321: Introduction to Deep Learning for Robotics

Neural Network N01: Multi-Input One-Output Model

10/16/2023

Outline

- Binary Classification
- Multi-Feature Input
- Binary Cross Entropy Loss
- Sigmoid Activation
- Gradient Descent

Multi-Input One-Output Model

Assume a dataset contains M objects

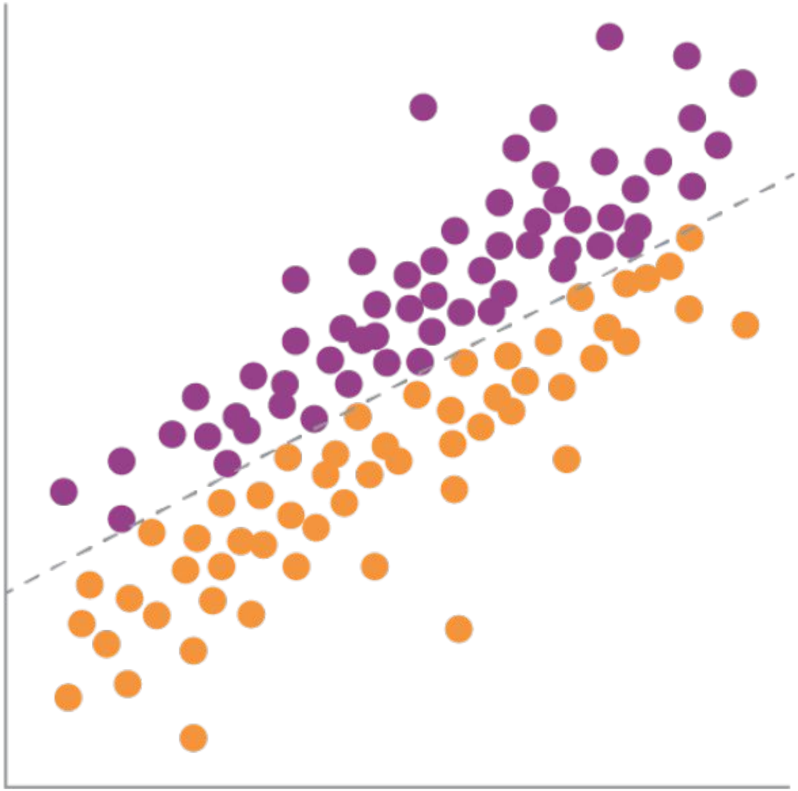
Each object can be described by multiple features: x_1 , x_2, ..., x_N

Each object has a known property/class: y

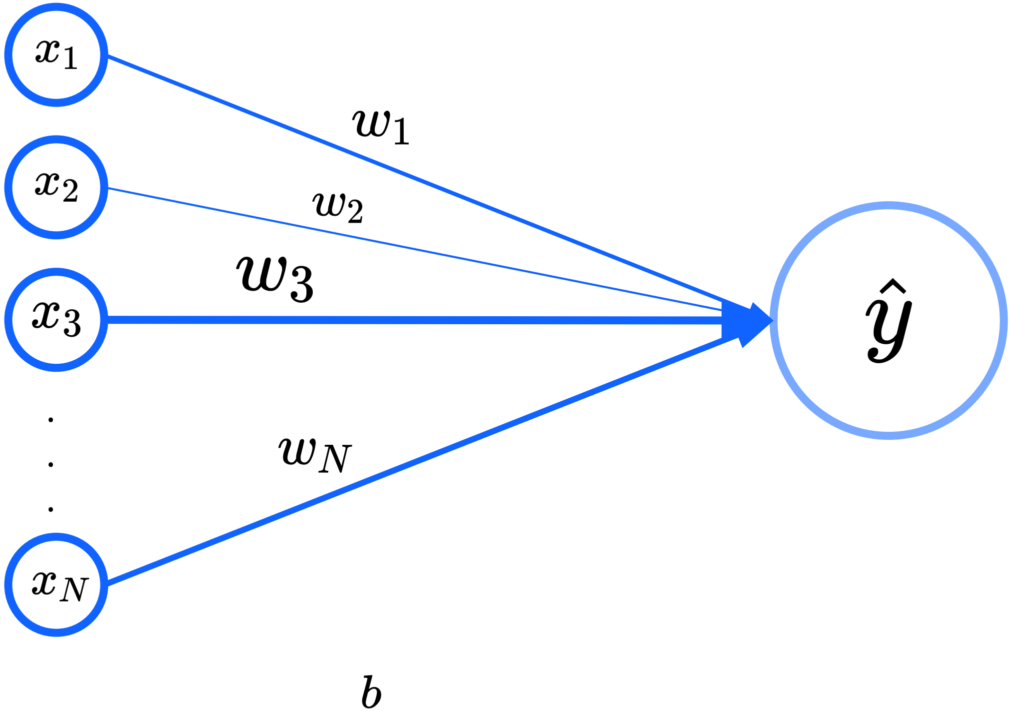
We would like to know the relationship between these features and the known property

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

Binary Classification



Neural Network Form



Matrix Form

$$
\begin{aligned}
\hat{\mathbf{y}} &= \sigma(\mathbf{Z}) \\
&= \sigma\left(\mathbf{X} \cdot \mathbf{w}^{\mathbf{T}} + \mathbf{b}\right)
\end{aligned}
$$

$$\sigma(\mathbf{Z}) = \frac{1}{1 + e^{(-\mathbf{Z})}}$$

Feature (Input) Matrix

$$\mathbf{x_n} = \begin{bmatrix} {}^{(1)}x_n \\ {}^{(2)}x_n \\ \cdot \\ \cdot \\ \cdot \\ {}^{(M)}x_n \end{bmatrix}_{(M,1)}$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{x_1} & \mathbf{x_2} & \dots & \mathbf{x_N} \end{bmatrix}_{(M,N)}$$

Target (Output) Vector

$$\mathbf{y} = \begin{bmatrix} {}^{(1)}y \\ {}^{(2)}y \\ . \\ . \\ . \\ {}^{(M)}y \end{bmatrix}_{(M,1)}$$

Model Parameters

$$\mathbf{w} = \begin{bmatrix} w_1 & w_2 & \dots & w_N \end{bmatrix}_{(1,N)}$$

$$\mathbf{b} = \begin{bmatrix} b \\ b \\ . \\ . \\ . \\ b \end{bmatrix}_{(M,1)}$$

Binary Cross Entropy Loss

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

Gradient of Loss

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

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \frac{\partial \mathcal{L}}{\partial \hat{\mathbf{y}}} \cdot \frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{Z}} \cdot \frac{\partial \mathbf{Z}}{\partial \mathbf{w}} = \frac{1}{M} \sum (\hat{\mathbf{y}} - \mathbf{y})^T \cdot \mathbf{X}$$

$$\frac{\partial \mathcal{L}}{\partial \mathrm{b}} = \frac{\partial \mathcal{L}}{\partial \hat{\mathbf{y}}} \cdot \frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{Z}} \cdot \frac{\partial \mathbf{Z}}{\partial \mathrm{b}} = \frac{1}{M} \sum (\hat{\mathbf{y}} - \mathbf{y})$$

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