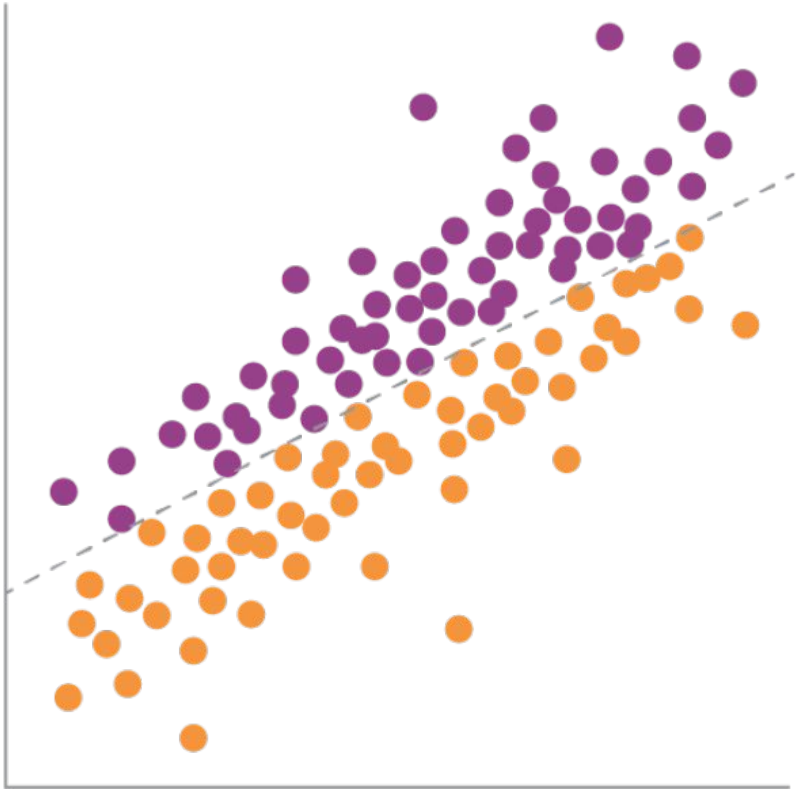
ENGR 3321: Introduction to Deep Learning for Robotics

Binary Classification

09/30/2024

Binary Classification



Binary Labels

A dataset with M samples:

$$\mathcal{D} = \{({}^{(1)}x_1, {}^{(1)}x_2, \ldots, {}^{(1)}x_N, {}^{(1)}y), ({}^{(2)}x_1, {}^{(2)}x_2, \ldots, {}^{(2)}x_N, {}^{(2)}y), \ldots, ({}^{(M)}x_1, {}^{(M)}x_2, \ldots, {}^{(M)}x_N, {}^{(M)}y)\}$$

$${}^{(i)}y \in \{0, 1\}$$

Feature (Input) Matrix

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

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

Model Parameters

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

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

Target (Output) Vector

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

Model in Matrix Form

$$\hat{\mathbf{y}} = \sigma(\mathbf{X} \cdot \mathbf{w}^T + \mathbf{b}) = \sigma(\mathbf{z})$$

$(M, 1)$ (M, N) $(M, 1)$ $(M, 1)$ $(M, 1)$

Matrix Multiplication

Binary Cross Entropy Loss

$$\mathcal{L}(\hat{\mathbf{y}}, \mathbf{y}) = \frac{1}{M}\sum_{i=1}^{M} -^{(i)}y \ln {}^{(i)}\hat{y} - (1 - {}^{(i)}y) \ln(1 - {}^{(i)}\hat{y}) = \overline{-\mathbf{y} \ln \hat{\mathbf{y}} - (1 - \mathbf{y}) \ln(1 - \hat{\mathbf{y}})}$$

Gradient of Loss

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

$$\frac{\partial \mathcal{L}}{\partial w_1} = \frac{\partial \mathcal{L}}{\partial \hat{\mathbf{y}}} \frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{z}} \frac{\partial \mathbf{z}}{\partial w_1} = \frac{1}{M} \sum_{i=1}^{M} ({}^{(i)}\hat{y} - {}^{(i)}y)^{(i)} x_1$$

$$\frac{\partial \mathcal{L}}{\partial w_2} = \frac{\partial \mathcal{L}}{\partial \hat{\mathbf{y}}} \frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{z}} \frac{\partial \mathbf{z}}{\partial w_2} = \frac{1}{M} \sum_{i=1}^{M} ({}^{(i)}\hat{y} - {}^{(i)}y)^{(i)} x_2$$

• • •

$$\frac{\partial \mathcal{L}}{\partial w_N} = \frac{\partial \mathcal{L}}{\partial \hat{\mathbf{y}}} \frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{z}} \frac{\partial \mathbf{z}}{\partial w_N} = \frac{1}{M} \sum_{i=1}^{M} ({}^{(i)}\hat{y} - {}^{(i)}y)^{(i)} x_N$$

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

Vectorized Gradient

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

$(1, N)$ $(1, M)$ (M, N)

Matrix Multiplication

$$\frac{\partial \mathcal{L}}{\partial b} = \overline{\hat{\mathbf{y}} - \mathbf{y}}$$

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