Machine Learning

4 – Implementing Logistic Regression

Avisek Gupta
Postdoctoral Fellow, IAI, TCG CREST
avisek003@gmail.com

August 23, 2022

Logistic Regression

We estimate the function:

$$\hat{y}_i = g\left(w_0 + w_1 x_{i1} + w_2 x_{i2} + \ldots + w_d x_{id}\right)$$

where,

$$g(t) = \frac{1}{1 + exp(-t)}$$

Logistic Regression

Why w_0 ?

We estimate the function:

$$\hat{y}_i = g(w_0 + w_1 x_{i1} + w_2 x_{i2} + \ldots + w_d x_{id})$$

where,

$$g(t) = \frac{1}{1 + exp(-t)}$$

Logistic Regression

Dropping w_0 limits the choice of hyperplanes to only those hyperplanes that pass through the origin

Why w_0 ?

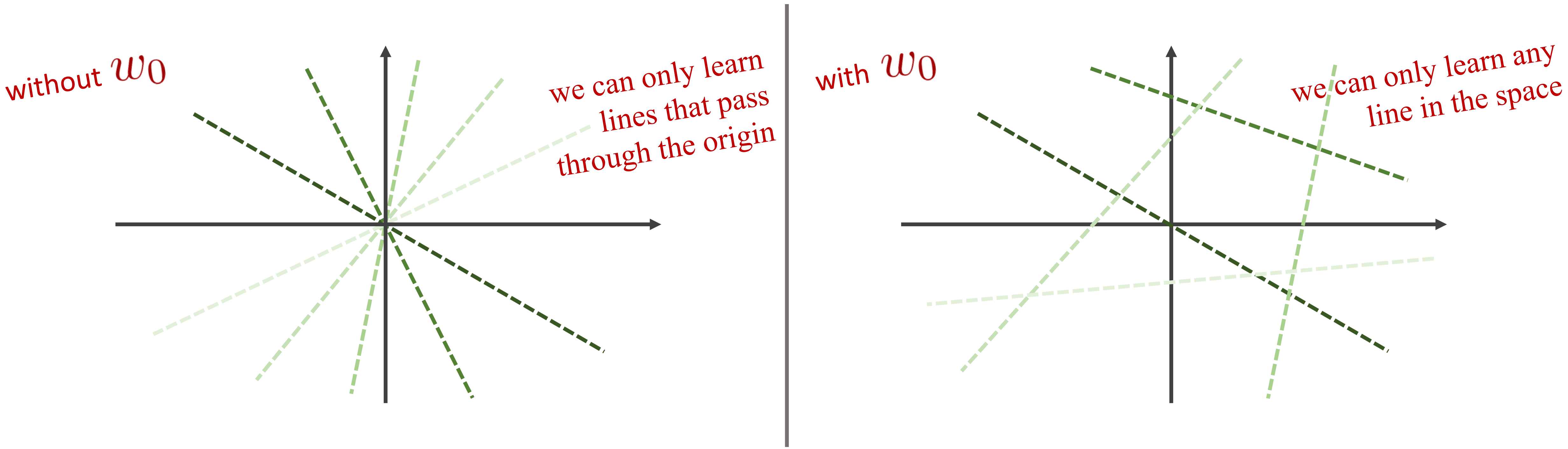
We estimate the function:

$$\hat{y}_i = g(w_0 + w_1 x_{i1} + w_2 x_{i2} + \ldots + w_d x_{id})$$

where,

$$g(t) = \frac{1}{1 + exp(-t)}$$

Examples in 2D:



Logistic Regression

We estimate the function:

$$\hat{y}_i = g(w_0 + w_1 x_{i1} + w_2 x_{i2} + \dots + w_d x_{id})$$

where,

$$g(t) = \frac{1}{1 + exp(-t)}$$

Choice of Loss Functions:

- Mean Square Error:

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

- Binary Cross Entropy Loss:

$$-\frac{1}{n}\sum_{i=1}^{n}\{y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\}$$

Using Gradient Descent on the MSE Loss Function

- We estimate the function:

$$\hat{y}_i = g(w_0 + w_1 x_{i1} + w_2 x_{i2} + \ldots + w_d x_{id})$$

where,

$$g(t) = \frac{1}{1 + exp(-t)}$$

- Mean Square Error Loss Function:

$$L_{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

- Gradient Descent Procedure:

1. Initialise $\mathbf{w}^{(0)}$
2. Update $\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - \eta \nabla_{\mathbf{w}^{(t)}} L_{MSE}$

Using Gradient Descent on the MSE Loss Function

- Rewriting with $x_{i0} = 1 \ \forall i$:

$$\hat{y}_i = g(w_0 x_{i0} + w_1 x_{i1} + w_2 x_{i2} + ... + +w_d x_{id}) = g(\mathbf{w}^T \mathbf{x}_i)$$

where,

$$g(t) = \frac{1}{1 + exp(-t)}$$

- The derivative of $g(t)$ is:

$$\frac{\partial}{\partial t} g(t) = \frac{(1+e^{-t})^2 \times 0 - 1 \times (-e^{-t})}{(1+e^{-t})^2} = \frac{e^{-t}}{(1+e^{-t})^2}$$

$$= \frac{1}{(1+e^{-t})} . \frac{e^{-t}}{(1+e^{-t})} = \frac{1}{(1+e^{-t})} . \left(1 - \frac{1}{(1+e^{-t})}\right)$$

$$= g(t).(1 - g(t))$$

Using Gradient Descent on the MSE Loss Function

- Logistic Regression:

$$\hat{y}_i = g(w_0 x_{i0} + w_1 x_{i1} + w_2 x_{i2} + \ldots + + w_d x_{id}) = g(\mathbf{w}^T \mathbf{x}_i)$$

where, $x_{i0} = 1 \ \forall i$, $g(t) = \frac{1}{1 + exp(-t)}$, and, $\frac{\partial}{\partial t} g(t) = g(t).(1 - g(t))$

- MSE Loss Function:

$$L_{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

- Gradient of the MSE Loss Function wrt w_j :

$$\nabla_{w_j} L_{MSE} = -\frac{2}{n} \sum_{i=1}^{n} \{(y_i - \hat{y}_i).g(\mathbf{w}^T \mathbf{x}_i).(1 - g(\mathbf{w}^T \mathbf{x}_i)).x_{ij}\}$$

$$= -\frac{2}{n} \sum_{i=1}^{n} \{(y_i - \hat{y}_i).\hat{y}_i(1 - \hat{y}_i).x_{ij}\}$$

Using Gradient Descent on the MSE Loss Function

- Logistic Regression:

$$\hat{y}_i = g(w_0 x_{i0} + w_1 x_{i1} + w_2 x_{i2} + \ldots + +w_d x_{id}) = g(\mathbf{w}^T \mathbf{x}_i)$$

where, $x_{i0} = 1 \ \forall i$, $g(t) = \frac{1}{1 + exp(-t)}$, and, $\frac{\partial}{\partial t} g(t) = g(t).(1 - g(t))$

- MSE Loss Function:

$$L_{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

- Gradient Descent to estimate w_j :

1. Initialise $w_j^{(0)}$
2. Update $w_j^{(t+1)} = w_j^{(t)} + \frac{2\eta}{n} \sum_{i=1}^{n} \{(y_i - \hat{y}_i).\hat{y}_i(1 - \hat{y}_i).x_{ij}\}$

Using Gradient Descent on the MSE Loss Function

Python Notebook: `LogisticRegression-MSE.ipynb`