Machine Learning Refined	Spring 2025
Topic: Regression — 3/1 2025	
FAUST	*Scribe: David Zhang*

1 Overview

In general, we aim to fit a line (or hyperplane in higher dimensions) to a scattering of data.

1.1 Notation and Modeling

Data for regression problems goes in the form of $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$ where $\mathbf{x}_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$.

Each input in $\mathbf{x}$ may be a column vector of length N.

Formally, the goal of regression is the following formula:

$$\underset{b,w}{\operatorname{argmin}} \sum_{p=1}^{N} (y_i - \mathbf{w}^T \mathbf{x}_i - b)^2 \tag{1}$$

where $\mathbf{w} \in \mathbb{R}^d$ is the weight vector and $b \in \mathbb{R}$ is the bias term.

The gradient of this cost after some chain rule:

$$\nabla g(w) = 2 \left(\sum_{p=1}^{N} x_p x_p^T \right) w - 2 \sum_{p=1}^{N} x_p y_p \tag{2}$$

Setting the gradient above to zero and solving for w gives the system of linear equations

$$\left(\sum_{p=1}^{N} x_p x_p^T \right) w = \sum_{p=1}^{N} x_p y_p \tag{3}$$

$$w^* = \left(\sum_{p=1}^{N} x_p x_p^T \right)^{-1} \sum_{p=1}^{N} x_p y_p \tag{4}$$

1.2 Efficacy of the Model

The efficacy of the model can be measured by the mean squared error (MSE) of the model:

$$\frac{1}{N}\sum_{p=1}^{N}(y_p - w^T x_p)^2 \tag{5}$$

However, as many functions are nonlinear, in order to fit a nonlinear function, we can use a nonlinear transformation, we can use an appropriate nonlinear feature transformation to describe nonlinear functions.

Problem 0. Model through a sin wave

Solution to Problem 0. Sinusoidal Regression One solution to this is to use a sinusoidal transformation. For example, we can use the following transformation:

$$b + f(x_p)w = b + sin(2\pi x_p)w \approx y_p, p = 1, \ldots, P \tag{6}$$

This transforms the feature space to $\{f(x_p), y_p\}_{p=1}^{P}$.

1.3 Computing $\mathbf{w}^*$

$$\underset{w}{\text{argmin}} = \sum_{p=1}^{P}(f_p w - y_p)^2$$

Deriving the derivative of the above equation gives us the following, setting to zero for the minimum:

$$2\sum_{p=1}^{P} f_p(f_p w - y_p) = 0$$

$$\sum_{p=1}^{P} f_p f_p w = \sum_{p=1}^{P} f_p y_p$$

$$w^* = \left(\sum_{p=1}^{P} f_p f_p^T\right)^{-1} \sum_{p=1}^{P} f_p y_p$$

Left on page 3.2.1

References

[1] Jeremy Wattós, Reza Borhanié Aggelos K. Katsaggelos, *Machine Learning Refined: Foundations, Algorithms, and Applications*, Northwestern University.