Linear Regression

Jerry Peng

1 Supervised & Unsupervised Learning

Supervised Learning The model is trained by learning the association between input data (features) and its corresponding output data (labels). In this training, the goal of the model is to "learn" the mapping from input to output by minimizing the gap between predicted values and actual values. Then, the model can be used to predict the output of new, unlabeled data. Common supervised learning tasks include classification (for example, determining whether an email is spam or not) and regression (for example, predicting house prices).

Unsupervised Learning The model only has input data (features) to learn from, without associated output labels. The goal of the model is to discover the structure or patterns in the input data. Common unsupervised learning tasks include clustering (for example, dividing customers into several different groups, with customers in each group having similar purchasing habits) and dimensionality reduction (for example, PCA).

How to tell? When determining whether a Machine Learning Algorithm is Supervised Learning, one needs to find out if the dataset used for training is in the form of $feature_i \rightarrow label_i$

2 Linear Regression

Regression Regression is a statistical and machine learning method used to understand and predict the relationship between two or more variables. In regression analysis, there is usually one dependent variable (also called the target variable or response variable) and one or more independent variables (also known as features or predictor variables). The goal of regression is to establish a model that describes the relationship between the dependent and independent variables. This model can be used to predict new, unknown data.

2.1 Definations

The simplest and most common type of regression is linear regression, which assumes a linear relationship between the dependent and independent variables. The form of the linear regression model is as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots \beta_n x_n + \epsilon \tag{1}$$

β is weight, ϵ is bias. The formula from above can be rewritten as:

$$y = wx + b \tag{2}$$

or:

$$X\theta = y \tag{3}$$

X is the set of the features, y is the set of the labels, θ is the set of the weights.

2.2 Solve for θ

We can use Ordinary Least Squares (OLS) to find the value of θ. The goal is to minimize the sum of squared residuals (SSR) between the predicted values of the model and the actual values. The detailed derivation is as follows: Suppose we have a Linear Regression, expressed as:

$$f(x) = X\theta + \epsilon \tag{4}$$

Then we can get SSR base on the definition:

$$\begin{aligned} SSR &= (y - X\theta)^T(y - X\theta) \\ &= (y^T - \theta^T X^T)(y - X\theta) \\ &= y^T y - y^T X\theta - \theta^T X^T y + \theta^T X^T X\theta \end{aligned} \tag{5}$$

Next, we find the derivative of SSR with respect to θ:

$$\begin{aligned} \frac{\partial SSR}{\partial \theta} &= 0 - X^T y - X^T y + 2X^T X\theta \\ &= -2X^T y + 2X^T X\theta \end{aligned} \tag{6}$$

Note: during the differentiation process, we use the following formula:

$$\frac{\partial (X^T X\theta)}{\partial \theta} = X^T X \tag{7}$$

$$\frac{\partial (y^T X\theta)}{\partial \theta} = X^T y \tag{8}$$

Setting the value of the derivative to zero, we get the following derivation:

$$\begin{aligned} \frac{\partial SSR}{\partial \theta} &= 0 \\ -2X^T y + 2X^T X\theta &= 0 \\ X^T y &= X^T X\theta \\ \theta &= (X^T X)^{-1} X^T y \end{aligned} \tag{9}$$

Then we can get the formula of θ:

$$\theta = (X^T X)^{-1} X^T y \tag{10}$$

References