Supervised Machine Learning: Regression and Classification

Rakshith Vaishnavi Dogra

Machine Learning

A field of study that allows computers to learn without being programmed explicitly

ML Algorithms

1. Supervised Learning Algorithms
2. Unsupervised Learning Algorithms
3. Recommendation Settings Algorithms
4. Reinforcement Learning Algorithms

Supervised Learning Algorithms

X → Y

Learns from Being given the Right Value of Y
• Regression - Predicting a number from infinitely many possible outcomes
• Classification - Predicting Category

Unsupervised Learning Algorithms

Algorithm has to find structures in data
• Clustering
• Dimensionality Reduction
• Anomaly Detection

Linear Regression Model

Terminology

x = input

y = output

m = number of Training examples

(x, y) = Simple Training Data

$(x^{(i)}, y^{(i)})$ = i^{th} training Example

$\hat{y}$ = estimate of y

$f_{w,b}(x) = wx + b$

$w \to$ weight

$b \to$ Bias

Squared Error Cost Function

$$J(w, b) = \frac{1}{2m} \sum_{i=1}^{m} (\hat{y}(i) - y(i))^2$$

$$J(w, b) = \frac{1}{2m} \sum_{\square i=1}^{\square m} (f_{w,b}(x^{(i)}) - y(i))^2$$

Gradient Descent for Linear Regression

For a Given $J(w, b)$, what is the $\text{Min}_{w,b} J(w, b)$

Gradient Descent Algorithm

$$w = w - \alpha \frac{\partial}{\partial w} J(w, b)$$

Where α is Learning Rate

$$b = b - \alpha \frac{\partial}{\partial w} J(w, b)$$

$$\frac{\partial}{\partial w} J(w, b) = \frac{1}{m} \sum_{i=1}^{m-1} (f_{w,b}(x^{(i)}) - y^{(i)}) x^{(i)}$$

$$\frac{\partial}{\partial b} J(w, b) = \frac{1}{m} \sum_{i=0}^{m-1} (f_{w,b}(x^{(i)}) - y^{(i)})$$

Batch Gradient Descent - Each step of Gradient Descent Uses all the Training Example

Multiple Regression

$x_j = j^{th}$ feature

n = Number of Features

$\vec{x}^{(i)}$ = Features of i^{th} training Example

$x^{(i)}{}_j$ = Value of Feature j in i^{th} training Example

$f_{\vec{w}, b}(X) = w_1 x_1 + w_2 x_2 + w_3 x_3 + \ldots \ldots w_n x_n$

$\vec{w} = [w_1, w_2, w_3, w_4, \ldots., w_n]$ (Row Vector)

$\vec{x} = [x_1, x_2, x_3, \ldots \ldots., x_n]$ (Row Vector)

$f_{\vec{w},b}(\vec{x}) = \vec{w}.\vec{x} + b$ (Multiple Linear Regression)

Polynomial Regression

$f_{\vec{w},b}(x) = w_1 x_1 + w_2 x_2 + b$

Classification

- Logistic Regression Model
- Overfitting - Regularization

Logistic Regression

$f_{\vec{w}, b} = \vec{w}.\vec{x} + b$

$$g(z) = \frac{1}{1 + g^{-z}}$$

$f_{\vec{w},b} = g(\vec{w}.\vec{x} + b)$

$$f_{\vec{w},b}(\vec{x}) = \frac{1}{1 + e^{-(\vec{w}.\vec{x}+b)}}$$

Cost Function for Logistic Regression

Logistic Loss Function

$$J(\vec{w}, b) = \frac{1}{m} \sum_{i=0}^{m-1} L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right)$$

Simplified Cost Function

$$L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}, y^{(i)}\right)\right) = -y^{(i)} \log\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) - (1 - y) \log\left(1 - f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right)$$

$$J(\vec{w}, b) = \frac{-1}{m} \sum_{i=1}^{m} \left[y^{(i)} \log\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) + \left(1 - y^{(i)}\right) \log\left(1 - f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right)\right]$$

Overfitting and Underfitting

Overfitting = High Variance

Under-fitting = High Bias

To Address Overfitting:-
- Collect More Data
- Reduce Parameters using Regularization
- Select features

Cost Function with Regularization

$$J\left(\vec{w}, b\right) = \frac{1}{2m}\sum_{i=1}^{m}\left(f_{\vec{w},b}\left(\vec{x}^{(x)}\right) - y^{(i)}\right)^2 + \frac{\lambda}{2m}\sum_{j=1}^{n} w_j^2 \text{ (Regularization Term)}$$

Gradient Descent with Regularized Linear Regression

$$w = \frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right) - y^{(i)}\right) x_j^{(i)} + \frac{\lambda}{m} w_j$$

$$w_j = w_j - \alpha\left\{\frac{1}{m}\sum_{i=1}^{m}\left[\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right) - y^{(i)}\right) x_j^{(i)}\right] + \frac{\lambda}{m} w_j\right\}$$

$$b = b - \alpha\left[\frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right) - y^{(i)}\right)\right]$$

$$\frac{\partial}{\partial w_j} J\left(\vec{w}, b\right) = \frac{1}{m}\sum_{i=1}^{m}\left[f_{\vec{w},b}\left(\vec{x}^{(i)} - y^{(i)}\right) x_j^{(i)}\right] + \frac{\lambda}{m} w_j$$

Regularized Linear Regression

$$J\left(\vec{w}, b\right) = \frac{-1}{m}\sum_{i=1}^{m}\left[y^{(i)} \log\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) + \left(1 - y^{(i)}\right) \log\left(1 - f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right)\right] + \frac{\lambda}{m}\sum_{j=1}^{n} w_j^2$$

$$w_j = \frac{1}{m}\sum_{i=1}^{m}\left[\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right) - y^{(i)}\right) x_j^{(i)}\right] + \frac{\lambda}{m} w_j$$