Regression Analysis

Dr. Chandranath Adak

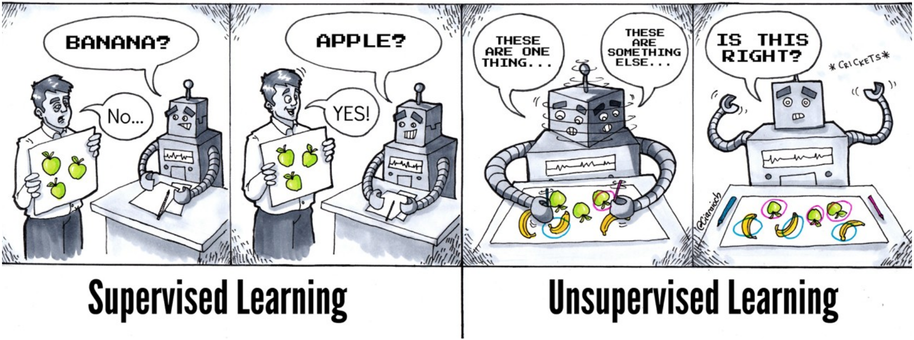
Dept. of CSE, Indian Institute of Technology Patna

September 1, 2025

Machine Learning

Learning

A computer program is said to learn from experience E with respect to some class of task T and performance measure P, if its performance at task in T as measured by P, improves with experience E.



Machine Learning: Label

What is regression

1. Supervised learning (we here look in terms of)
2. Estimating the relationships between a dependent variable (also termed as outcome) and one or more independent variables (often termed as features)

What is regression

1. Supervised learning (we here look in terms of)
2. Estimating the relationships between a dependent variable (also termed as outcome) and one or more independent variables (often termed as features)
3. In regression, the aim is to find the relation (often termed as a hypothesis) h_θ between some input variables $X \in \mathbb{R}^k$ and an output variable $y \in \mathbb{R}$

What is regression

1. Supervised learning (we here look in terms of)
2. Estimating the relationships between a dependent variable (also termed as outcome) and one or more independent variables (often termed as features)
3. In regression, the aim is to find the relation (often termed as a hypothesis) h_θ between some input variables $X \in \mathbb{R}^k$ and an output variable $y \in \mathbb{R}$

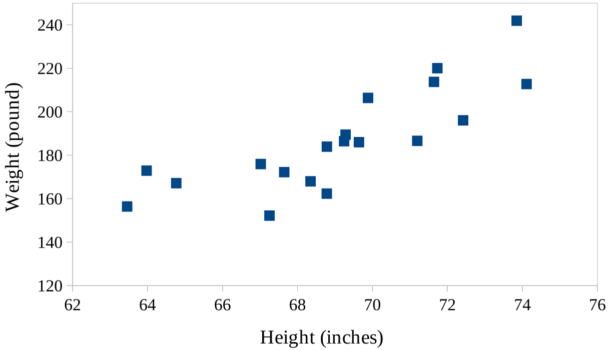
Examples

- Features: {Outside temperature, People inside room, target room temperature};
  Outcome: Energy requirement
- Features: {Size, Number of Bedrooms, Number of Floors, Age of the Home};
  Outcome: Price

Linear regression with single variable

- We have a single independent variable and a dependent variable
- Scatter plot indicates a linear relationship between independent and dependent variables

Height	Weight
63	156
64	173
65	167
67	176
67	152
68	172
68	168
69	162
69	184
69	186
69	189
70	186
70	206
71	187
72	214
72	220
72	196
74	242
74	213



Mathematical formulation of linear regression with single variable

- Given n observations $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, the aim is to find the linear relationship / hypothesis h_θ

Mathematical formulation of linear regression with single variable

- Given n observations $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, the aim is to find the linear relationship / hypothesis h_θ
- The linear regression with a single variable model is:

$$y = \theta_0 + \theta_1 x + \epsilon$$

  - x : Regressor variable
  - y : Response variable
  - θ_0 : Intercept
  - θ_1 : Slope
  - ϵ : Random error

Mathematical formulation of linear regression with single variable

- Given n observations $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, the aim is to find the linear relationship / hypothesis h_θ
- The linear regression with a single variable model is:

$$y = \theta_0 + \theta_1 x + \epsilon$$

  - x : Regressor variable
  - y : Response variable
  - θ_0 : Intercept
  - θ_1 : Slope
  - ϵ : Random error
- Therefore, $\forall\, i \in \{1, 2, \ldots, n\};$ we have, $y_i = \theta_0 + \theta_1 x_i + \epsilon_i$
- $\hat{y}_i = h_\theta(x_i) = \theta_0 + \theta_1 x_i$,
  where $\hat{y}_i$ is the predicted value under hypothesis h_θ
- For a given x_i, the corresponding observation y_i has the value $\theta_0 + \theta_1 x_i$ plus an error component

Mathematical formulation (Contd.)

Assumption on the error component

For $y_i = \theta_0 + \theta_1 x_i + \epsilon_i, \quad i = \{1, 2, \ldots, n\}$

1. ϵ_i is a normally distributed random variable with mean 0 and variance σ^2, i.e., $\epsilon_i \sim N(0, \sigma^2)$
2. ϵ_i and ϵ_j are uncorrelated, $i \neq j$, i.e., $Cov(\epsilon_i, \epsilon_j) = 0$

Mathematical formulation (Contd.)

Assumption on the error component

For $y_i = \theta_0 + \theta_1 x_i + \epsilon_i, \quad i = \{1, 2, \ldots, n\}$

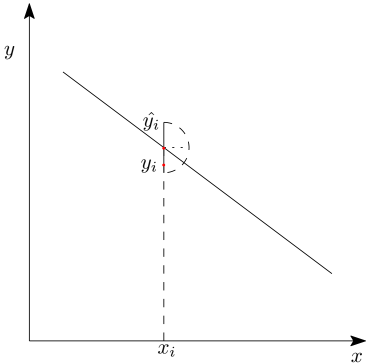
1. ϵ_i is a normally distributed random variable with mean 0 and variance σ^2, i.e., $\epsilon_i \sim N(0, \sigma^2)$
2. ϵ_i and ϵ_j are uncorrelated, $i \neq j$, i.e., $Cov(\epsilon_i, \epsilon_j) = 0$

- $E(y_i) = E(\theta_0 + \theta_1 x_i + \epsilon_i) = E(\theta_0 + \theta_1 x_i) \quad$ (as $E(\epsilon_i) = 0$)
- $Var(y_i) = Var(\theta_0 + \theta_1 x_i + \epsilon_i) = Var(\epsilon_i) = \sigma^2$

Consequence

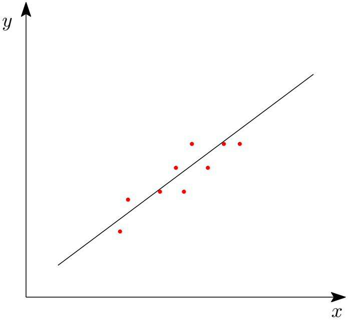
$$y_i \sim N(\theta_0 + \theta_1 x_i, \sigma^2)$$

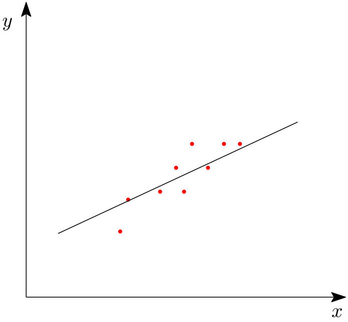
Mathematical formulation (Contd.)



$$y_i \sim N(\theta_0 + \theta_1 x_i, \sigma^2)$$

Solution to regression problem





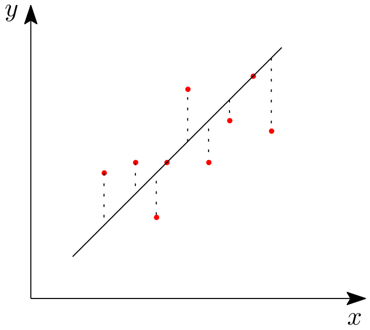
What should be the curve?

One possible solution

Least square method

Error

- Error $e_i = (y_i - \hat{y}_i), \; \forall i \in \{1, 2, \ldots, n\}$



Objective

To estimate θ_0 and θ_1 to minimize

$$SS_{Res} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 x_i)^2$$

R^2

R-Squared (Coefficient of determination)

$$R^2 = 1 - \frac{SS_{Res}}{SS_{Total}} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \mu_y)^2} \tag{1}$$

R^2

R-Squared (Coefficient of determination)

$$R^2 = 1 - \frac{SS_{Res}}{SS_{Total}} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \mu_y)^2} \tag{1}$$

$R^2 = 1$ (best)
$R^2 = 0$ (worst)

Least square method: Normal equations

$$S = \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 x_i)^2$$

$$\frac{\partial S}{\partial \theta_0}\Big|_{\hat{\theta}_0, \hat{\theta}_1} = -2 \sum_{i=1}^{n} (y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i) = 0 \tag{2}$$

$$\frac{\partial S}{\partial \theta_1}\Big|_{\hat{\theta}_0, \hat{\theta}_1} = -2 \sum_{i=1}^{n} (y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i) x_i = 0 \tag{3}$$

Least square method: Normal equations

$$
\begin{aligned}
\frac{\partial S}{\partial \theta_0}\Big|_{\hat{\theta}_0,\hat{\theta}_1} = \quad -2\sum_{i=1}^{n}(y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i) &= 0 \\
\sum_{i=1}^{n}(y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i) &= 0
\end{aligned}
$$

Least square method: Normal equations

$$\begin{aligned}
\frac{\partial S}{\partial \theta_0}\Big|_{\hat{\theta}_0,\hat{\theta}_1} = \quad -2\sum_{i=1}^{n}(y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i) &= 0 \\
\sum_{i=1}^{n}(y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i) &= 0 \\
\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \hat{\theta}_0 - \hat{\theta}_1 \sum_{i=1}^{n} x_i &= 0
\end{aligned}$$

Least square method: Normal equations

$$
\begin{aligned}
\frac{\partial S}{\partial \theta_0}\Big|_{\hat{\theta}_0,\hat{\theta}_1} = \quad -2\sum_{i=1}^{n}(y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i) &= 0 \\
\sum_{i=1}^{n}(y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i) &= 0 \\
\sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\hat{\theta}_0 - \hat{\theta}_1 \sum_{i=1}^{n} x_i &= 0 \\
n\mu_y - n\hat{\theta}_0 - \hat{\theta}_1 n\mu_x &= 0
\end{aligned}
$$

Least square method: Normal equations

$$
\begin{aligned}
\frac{\partial S}{\partial \theta_0}\Big|_{\hat{\theta}_0,\hat{\theta}_1} = \quad -2\sum_{i=1}^{n}(y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i) &= 0 \\
\sum_{i=1}^{n}(y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i) &= 0 \\
\sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\hat{\theta}_0 - \hat{\theta}_1 \sum_{i=1}^{n} x_i &= 0 \\
n\mu_y - n\hat{\theta}_0 - \hat{\theta}_1 n\mu_x &= 0
\end{aligned}
$$

$$\hat{\theta}_0 = \mu_y - \hat{\theta}_1 \mu_x$$

Least square method: Normal equations

$$
\begin{aligned}
\frac{\partial S}{\partial \theta_1}\Big|_{\hat{\theta}_0,\hat{\theta}_1} = \quad & -2\sum_{i=1}^{n}(y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i)x_i && = 0 \\
& \sum_{i=1}^{n}(y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i)x_i && = 0
\end{aligned}
$$

Least square method: Normal equations

$$
\begin{aligned}
\frac{\partial S}{\partial \theta_1}\Big|_{\hat{\theta}_0,\hat{\theta}_1} = \quad -2\sum_{i=1}^{n}(y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i)x_i &= 0 \\
\sum_{i=1}^{n}(y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i)x_i &= 0 \\
\sum_{i=1}^{n}(y_i - \mu_y + \hat{\theta}_1\mu_x - \hat{\theta}_1 x_i)x_i &= 0
\end{aligned}
$$

Least square method: Normal equations

$$
\begin{aligned}
\frac{\partial S}{\partial \theta_1}\Big|_{\hat{\theta}_0,\hat{\theta}_1} = \quad -2\sum_{i=1}^{n}(y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i)x_i &= 0 \\
\sum_{i=1}^{n}(y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i)x_i &= 0 \\
\sum_{i=1}^{n}(y_i - \mu_y + \hat{\theta}_1\mu_x - \hat{\theta}_1 x_i)x_i &= 0 \\
\sum_{i=1}^{n}(y_i - \mu_y)x_i - \hat{\theta}_1 \sum_{i=1}^{n}(x_i - \mu_x)x_i &= 0
\end{aligned}
$$

Least square method: Normal equations

$$
\begin{aligned}
\frac{\partial S}{\partial \theta_1}\Big|_{\hat{\theta}_0,\hat{\theta}_1} = \quad & -2\sum_{i=1}^{n}(y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i)x_i & = 0 \\
& \sum_{i=1}^{n}(y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i)x_i & = 0 \\
& \sum_{i=1}^{n}(y_i - \mu_y + \hat{\theta}_1 \mu_x - \hat{\theta}_1 x_i)x_i & = 0 \\
& \sum_{i=1}^{n}(y_i - \mu_y)x_i - \hat{\theta}_1 \sum_{i=1}^{n}(x_i - \mu_x)x_i & = 0 \\
\hat{\theta}_1 = \quad & \frac{\sum_{i=1}^{n}(y_i - \mu_y)x_i}{\sum_{i=1}^{n}(x_i - \mu_x)x_i} &
\end{aligned}
$$

Least square method: Normal equations

$$
\begin{aligned}
\frac{\partial S}{\partial \theta_1}\Big|_{\hat{\theta}_0,\hat{\theta}_1} = \quad & -2\sum_{i=1}^{n}(y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i)x_i &= 0\\
& \sum_{i=1}^{n}(y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i)x_i &= 0\\
& \sum_{i=1}^{n}(y_i - \mu_y + \hat{\theta}_1\mu_x - \hat{\theta}_1 x_i)x_i &= 0\\
& \sum_{i=1}^{n}(y_i - \mu_y)x_i - \hat{\theta}_1 \sum_{i=1}^{n}(x_i - \mu_x)x_i &= 0\\
\hat{\theta}_1 = \quad & \frac{\sum_{i=1}^{n}(y_i - \mu_y)x_i}{\sum_{i=1}^{n}(x_i - \mu_x)x_i} &\\
\hat{\theta}_1 = \quad & \frac{\sum_{i=1}^{n}(y_i - \mu_y)(x_i - \mu_x)}{\sum_{i=1}^{n}(x_i - \mu_x)^2} &
\end{aligned}
$$

Least square method: Normal equations

$$\begin{aligned}
\frac{\partial S}{\partial \theta_1}\Big|_{\hat{\theta}_0,\hat{\theta}_1} = \quad & -2\sum_{i=1}^{n}(y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i)x_i & = 0 \\
& \sum_{i=1}^{n}(y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i)x_i & = 0 \\
& \sum_{i=1}^{n}(y_i - \mu_y + \hat{\theta}_1\mu_x - \hat{\theta}_1 x_i)x_i & = 0 \\
& \sum_{i=1}^{n}(y_i - \mu_y)x_i - \hat{\theta}_1 \sum_{i=1}^{n}(x_i - \mu_x)x_i & = 0 \\
\hat{\theta}_1 = \quad & \frac{\sum_{i=1}^{n}(y_i - \mu_y)x_i}{\sum_{i=1}^{n}(x_i - \mu_x)x_i} & \\
\hat{\theta}_1 = \quad & \frac{\sum_{i=1}^{n}(y_i - \mu_y)(x_i - \mu_x)}{\sum_{i=1}^{n}(x_i - \mu_x)^2} &
\end{aligned}$$

$$\hat{\theta}_1 = \frac{Cov(X,Y)}{Var(X)}$$

Mathematical formulation of linear regression with multiple variables

- Given n observations
  $\{(x_1^1, x_2^1, \ldots, x_k^1, y^1), (x_1^2, x_2^2, \ldots, x_k^2, y^2), \ldots, (x_1^n, x_2^n, \ldots, x_k^n, y^n)\}$
  Aim is to find the linear relationship / hypothesis h_θ
- The linear regression with k variables model is:

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k + \epsilon$$

  - k Regressor variables $x_1, x_2, \ldots, x_k$
  - y be the response variable

Mathematical formulation of linear regression with multiple variables

- Given n observations $\{(x_1^1, x_2^1, \ldots, x_k^1, y^1), (x_1^2, x_2^2, \ldots, x_k^2, y^2), \ldots, (x_1^n, x_2^n, \ldots, x_k^n, y^n)\}$
  Aim is to find the linear relationship / hypothesis h_θ
- The linear regression with k variables model is:

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k + \epsilon$$

  - k Regressor variables $x_1, x_2, \ldots, x_k$
  - y be the response variable
- We have $(k+1)$ normal equations considering the least square loss

$$\frac{\partial S}{\partial \theta_j}\Big|_{\hat{\theta}_0, \hat{\theta}_1, \ldots, \hat{\theta}_k} = -2 \sum_{i=1}^{n} (y^i - \hat{\theta}_0 - \hat{\theta}_1 x_1^i - \hat{\theta}_2 x_2^i - \cdots - \hat{\theta}_k x_k^i) x_j^i = 0 \qquad (4)$$

provided $\forall i \in \{1, 2, \ldots, n\}, \qquad x_0^i = 1$

Linear regression with multiple variables

- Feature matrix with $n \times (k+1)$ dimension

$$X = \begin{bmatrix} 1 & x_1^1 & x_2^1 & \dots & x_k^1 \\ 1 & x_1^2 & x_2^2 & \dots & x_k^2 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_1^n & x_2^n & \dots & x_k^n \end{bmatrix}$$

- Outcome matrix with $n \times 1$ dimension $Y = \begin{bmatrix} y^1 \\ y^2 \\ \dots \\ y^n \end{bmatrix}$

- Parameter matrix with $(k+1) \times 1$ dimension $\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \dots \\ \theta_k \end{bmatrix}$

$$\hat{\theta} = (X^T X)^{-1} X^T Y$$

Linear regression with multiple variables

$\hat{Y} = X\theta$

$S = \frac{1}{n}(X\theta - Y)^T(X\theta - Y)$

$S = \frac{1}{n}(\theta^T X^T - Y^T)(X\theta - Y)$

Linear regression with multiple variables

$$\hat{Y} = X\theta$$

$$S = \frac{1}{n}(X\theta - Y)^T(X\theta - Y)$$

$$S = \frac{1}{n}(\theta^T X^T - Y^T)(X\theta - Y)$$

$$S = \frac{1}{n}(\theta^T X^T X\theta - Y^T X\theta - \theta^T X^T Y + Y^T Y)$$

Linear regression with multiple variables

$$\hat{Y} = X\theta$$

$$S = \frac{1}{n}(X\theta - Y)^T(X\theta - Y)$$

$$S = \frac{1}{n}(\theta^T X^T - Y^T)(X\theta - Y)$$

$$S = \frac{1}{n}(\theta^T X^T X\theta - Y^T X\theta - \theta^T X^T Y + Y^T Y)$$

$$S = \frac{1}{n}(\theta^T X^T X\theta - 2\theta^T X^T Y + Y^T Y)$$

Linear regression with multiple variables

$\hat{Y} = X\theta$

$S = \frac{1}{n}(X\theta - Y)^T(X\theta - Y)$

$S = \frac{1}{n}(\theta^T X^T - Y^T)(X\theta - Y)$

$S = \frac{1}{n}(\theta^T X^T X\theta - Y^T X\theta - \theta^T X^T Y + Y^T Y)$

$S = \frac{1}{n}(\theta^T X^T X\theta - 2\theta^T X^T Y + Y^T Y)$

$\frac{\partial S}{\partial \theta} = \frac{1}{n}(2X^T X\theta - 2X^T Y)$

Linear regression with multiple variables

$\hat{Y} = X\theta$

$S = \frac{1}{n}(X\theta - Y)^T(X\theta - Y)$

$S = \frac{1}{n}(\theta^T X^T - Y^T)(X\theta - Y)$

$S = \frac{1}{n}(\theta^T X^T X\theta - Y^T X\theta - \theta^T X^T Y + Y^T Y)$

$S = \frac{1}{n}(\theta^T X^T X\theta - 2\theta^T X^T Y + Y^T Y)$

$\frac{\partial S}{\partial \theta} = \frac{1}{n}(2X^T X\theta - 2X^T Y)$

Normal equation: $\frac{\partial S}{\partial \theta} = \frac{1}{n}(2X^T X\theta - 2X^T Y) = 0$

$(2X^T X\theta - 2X^T Y) = 0$

$X^T X\theta = X^T Y$

$\theta = (X^T X)^{-1} X^T Y$

Linear regression with multiple variables

- Linear regression can be solved deterministically using normal equations
- The complexity to find out $(X^T X)^{-1}$ is $O(k^3)$, k is the number of features
- For large k, this solution is not scalable

Can you do something else?

Gradient descent

- Gradient descent is a first-order iterative optimization algorithm for finding a local minimum of a differentiable function
- Consider a function $J(\theta_0, \theta_1)$
- Objective is to find $\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$

Gradient descent

- Gradient descent is a first-order iterative optimization algorithm for finding a local minimum of a differentiable function
- Consider a function $J(\theta_0, \theta_1)$
- Objective is to find $\min_{\theta_0,\theta_1} J(\theta_0, \theta_1)$

Algorithm

1. Start with any arbitrary values of θ_0 and θ_1
2. In each iteration, update the values of θ_0 and θ_1 to reduce the value of $J(\theta_0, \theta_1)$
3. Terminate the algorithm once *termination criteria* is satisfied

Gradient descent algorithm

Repeat until terminated$\{$

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \qquad \text{(for } j = \{0, 1\}\text{)}$$

$\}$

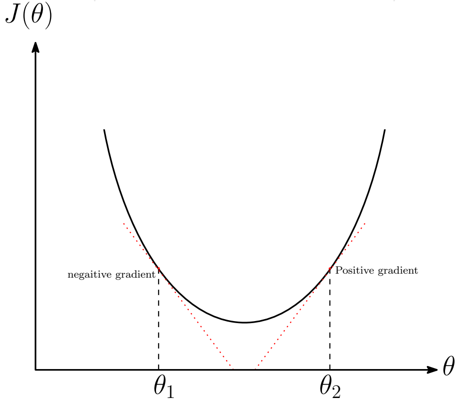
- α is a hyperparameter representing the learning rate

Why derivatives in gradient descent?

Repeat until terminated{

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \qquad \text{(for } j = \{0, 1\}\text{)}$$

}



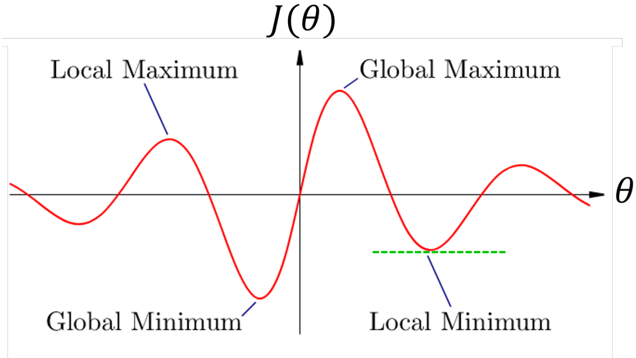
Why derivatives in gradient descent?

Repeat until terminated{

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \qquad \text{(for } j = \{0, 1\})$$

}

- What happens if the local minima is reached?



Why learning rate in gradient descent?

Repeat until terminated{

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \qquad \text{(for } j = \{0, 1\}\text{)}$$

}

- What happens if the value of α is very small?

Why learning rate in gradient descent?

Repeat until terminated{

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \qquad \text{(for } j = \{0, 1\}\text{)}$$

}

- What happens if the value of α is very small?
  - Learning will be done slowly
- What happens if the value of α is very big?

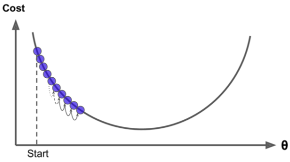
Why learning rate in gradient descent?

Repeat until terminated{

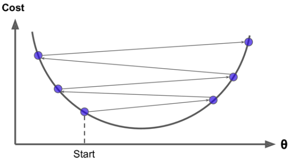
$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \qquad (\text{for } j = \{0, 1\})$$

}

- What happens if the value of α is very small?
  - Learning will be done slowly
- What happens if the value of α is very big?
  - May overshoot the minimum value
  - May fail to converge, even diverge



a) too small



a) too big

Linear regression using gradient descent

- The purpose of the gradient descent is to optimize the cost function associated with linear regression
- Cost function: $J_\theta = \frac{1}{2n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$
  where $\hat{y}_i = h_\theta(x_i) = \theta_0 + \theta_1 x_i$ (For single variable linear regression)
- The aim of gradient descent is to minimize $J(\theta_0, \theta_1)$
  $\frac{\partial}{\partial\theta_0}J(\theta_0,\theta_1) = \frac{\partial}{\partial\theta_0}\frac{1}{2n}\sum_{i=1}^{n}(\theta_0 + \theta_1 x_i - y_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(\theta_0 + \theta_1 x_i - y_i)$
  $\frac{\partial}{\partial\theta_1}J(\theta_0,\theta_1) = \frac{\partial}{\partial\theta_1}\frac{1}{2n}\sum_{i=1}^{n}(\theta_0 + \theta_1 x_i - y_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(\theta_0 + \theta_1 x_i - y_i)x_i$

For single variable:

Repeat until terminated{

$$\theta_0 = \theta_0 - \alpha \frac{1}{n} \sum_{i=1}^{n} (\theta_0 + \theta_1 x_i - y_i)$$

$$\theta_1 = \theta_1 - \alpha \frac{1}{n} \sum_{i=1}^{n} (\theta_0 + \theta_1 x_i - y_i) x_i$$

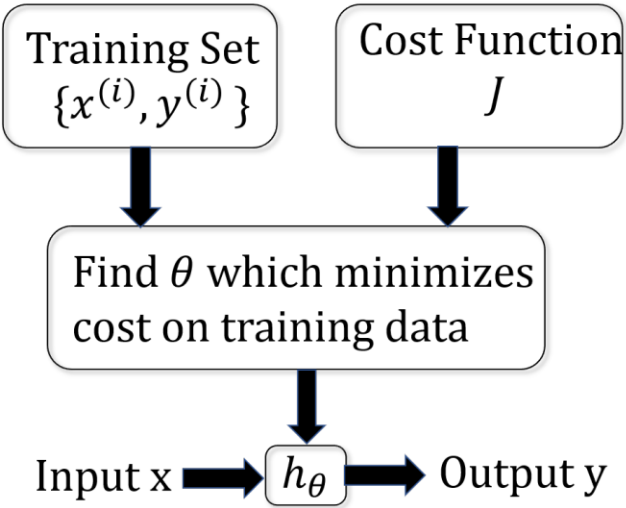
}

For multiple variables:

Repeat until terminated{

$\forall j \in \{0, 1, 2, \ldots, k\}$ $$\theta_j = \theta_j - \alpha \frac{1}{n} \sum_{i=1}^{n} (\theta_0 + \theta_1 x_j^i - y^i) x_j^i$$

}


Training Set
$\{x^{(i)}, y^{(i)}\}$
Cost Function
J
Find θ which minimizes cost on training data
Input x
h_θ
Output y

Questions

- How do you select the value of α?

Questions

- How do you select the value of α?
- What should be the termination criteria of gradient descent?

Questions

- How do you select the value of α?
- What should be the termination criteria of gradient descent?
- Is fixed value of α sufficient to converge the gradient descent?

Questions

- How do you select the value of α?
- What should be the termination criteria of gradient descent?
- Is fixed value of α sufficient to converge the gradient descent?
- Feature scaling in gradient descent

Questions

- How do you select the value of α?
- What should be the termination criteria of gradient descent?
- Is fixed value of α sufficient to converge the gradient descent?
- Feature scaling in gradient descent
- Comparison between gradient descent and normal equation methods.

Questions

- How do you select the value of α?
- What should be the termination criteria of gradient descent?
- Is fixed value of α sufficient to converge the gradient descent?
- Feature scaling in gradient descent
- Comparison between gradient descent and normal equation methods.
- Polynomial regression

Questions

- How do you select the value of α?
- What should be the termination criteria of gradient descent?
- Is fixed value of α sufficient to converge the gradient descent?
- Feature scaling in gradient descent
- Comparison between gradient descent and normal equation methods.
- Polynomial regression
- Different types of gradient descent techniques

Thank You!