Linear Regression

Pritam Prakash Shete
Computer Division, BARC
Centre for Excellence in Basic Sciences

Topics

- Supervised learning – Regression model
- Linear regression
- Least squares
- R^2
- Adjusted R^2
- Applications
- Advantages
- Disadvantages

Supervised Learning

- Training set – { $(x^{(1)},y^{(1)})$, $(x^{(2)},y^{(2)})$, …, $(x^{(m)},y^{(m)})$}
- Labeled dataset

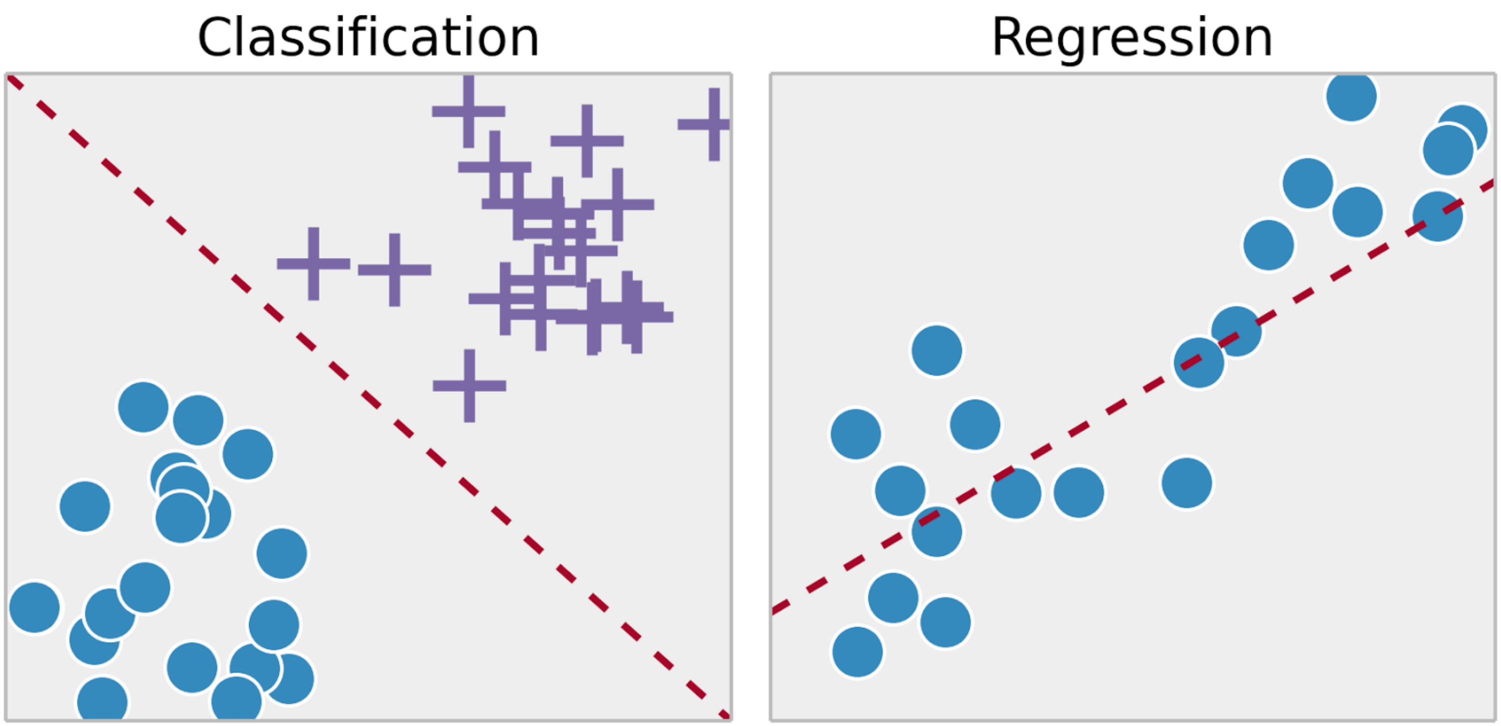


Image ack – Google

Linear Regression

- Independent variable – X axis
- Dependent variable – Y axis
- Data points – Samples
- Relationship
- Line of regression

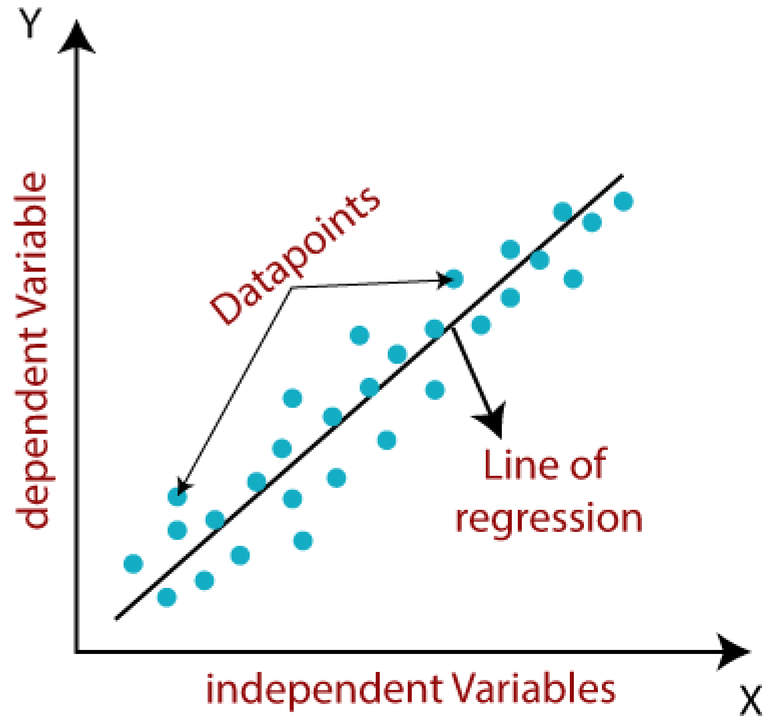


Image ack – Google

Linear Regression

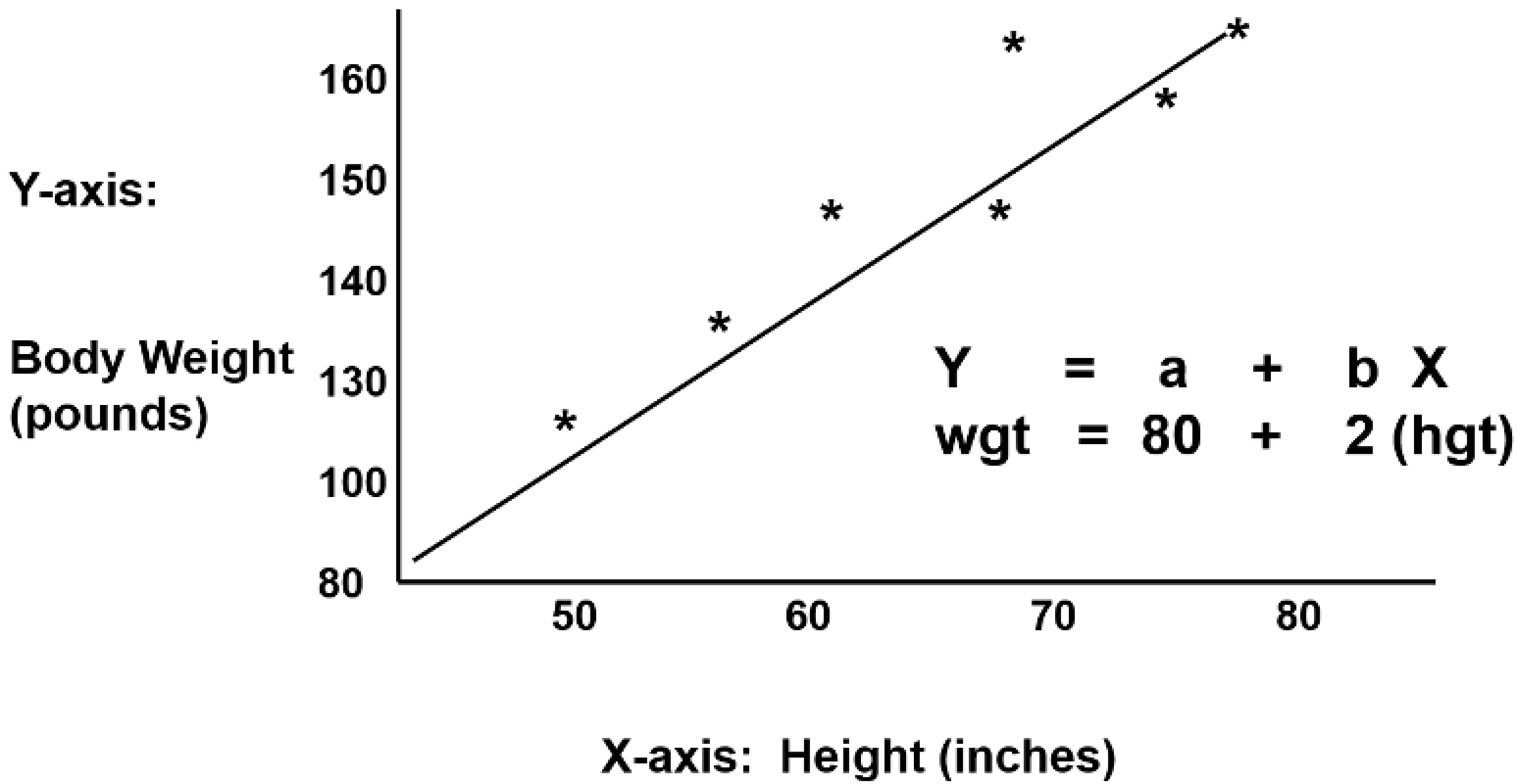


Image ack – Google

Linear Regression

- Input – X
  - Vector
  - X ε R
  - Dimension – n_X
- Output – y
  - Scalar
  - y ε R

Linear Regression

- Weights – W – (W_1, W_2, W_3, …, W_{nx})
  - Vector
  - W ε R
  - Dimension – n_X
- Bias – b – W_0
  - Scalar
  - b ε R
- $y = W^T X + b$

Linear Regression

- Positive correlation
  - Independent variable increases
  - Dependent variable increases
- Negative correlation
  - Independent variable increases
  - Dependent variable decreases
- No correlation
  - Independent variable increases
  - Dependent variable no change

Linear Regression

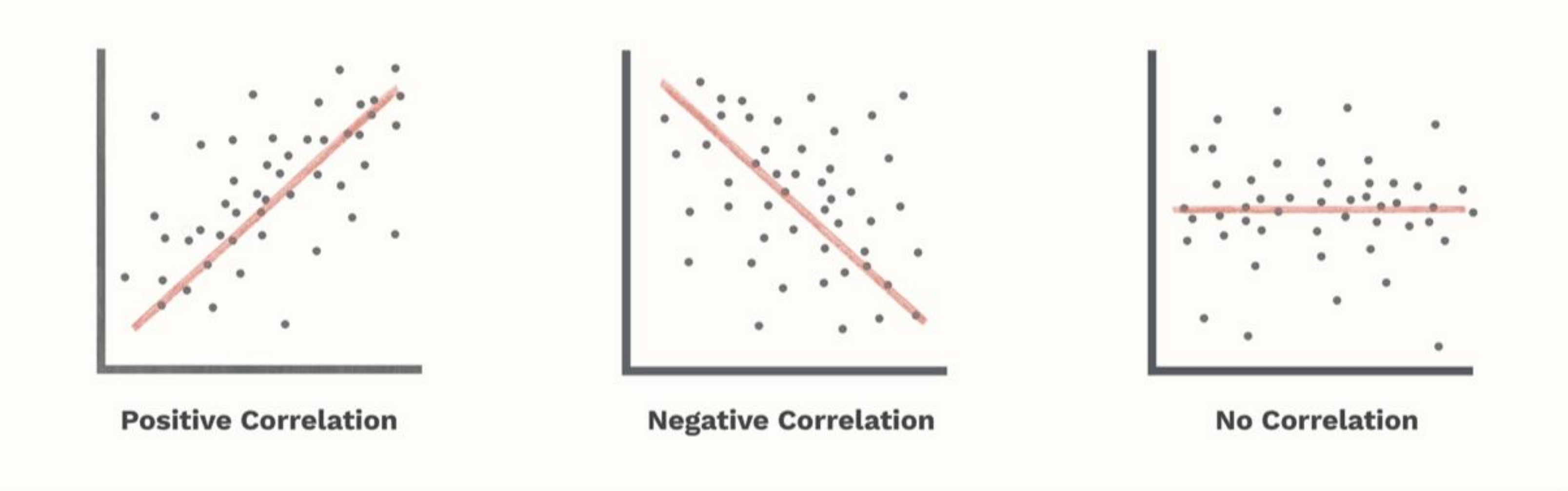


Image ack – Google

Linear Regression – Types

- Simple linear regression
  – Single independent variable
- Multiple linear regression
  – More than one independent variable

Least Squares

- Linear regression – Least squares
- Actual value – Observed value
- Estimated value – Predicted value
- Error value
- Sum of squared residuals
- Minimize error value
- Generic equation of line

Least Squares

$$\widehat{y}_i = W_0 + W_1 \times X_i$$

$$y_i = \widehat{y}_i + \varepsilon_i$$

$$\mathrm{J} = \frac{1}{\mathrm{n_x}} \sum_{i=1}^{n_x} \varepsilon_i^{\ 2}$$

- X_i – Independent variable
- y_i – Dependent variable
- W_0 – Y axis intercept
- W_1 – Slope of line
- $\hat{y}_i$ – Estimated value
- ε_i – Random error
- Mean squared error – MSE
- J – Cost function – Minimize cost function

Loss Function

- Loss function
  - One sample – i^{th} sample
  - $L(y^{(i)}, \hat{y}^{(i)}) = (y^{(i)} - \hat{y}^{(i)})^2$
- Cost function
  - Average of loss function for all samples

$$J(W,b) = \frac{1}{\mathrm{n_x}} \sum_{i=1}^{n_x} L(y^{(i)}, \hat{y}^{(i)})$$

$$J(W,b) = \frac{1}{\mathrm{n_x}} \sum_{i=1}^{n_x} \left(y^{(i)} - \hat{y}^{(i)}\right)^2$$

Gradient Descent

- Input dataset – $\{ (x^{(1)},y^{(1)}), (x^{(2)},y^{(2)}), \ldots, (x^{(m)},y^{(m)}) \}$
- Equations – $\hat{y} = W^T X + b$
- Loss function – $L(\hat{y}, y) = (y - \hat{y})^2$
- Cost function –

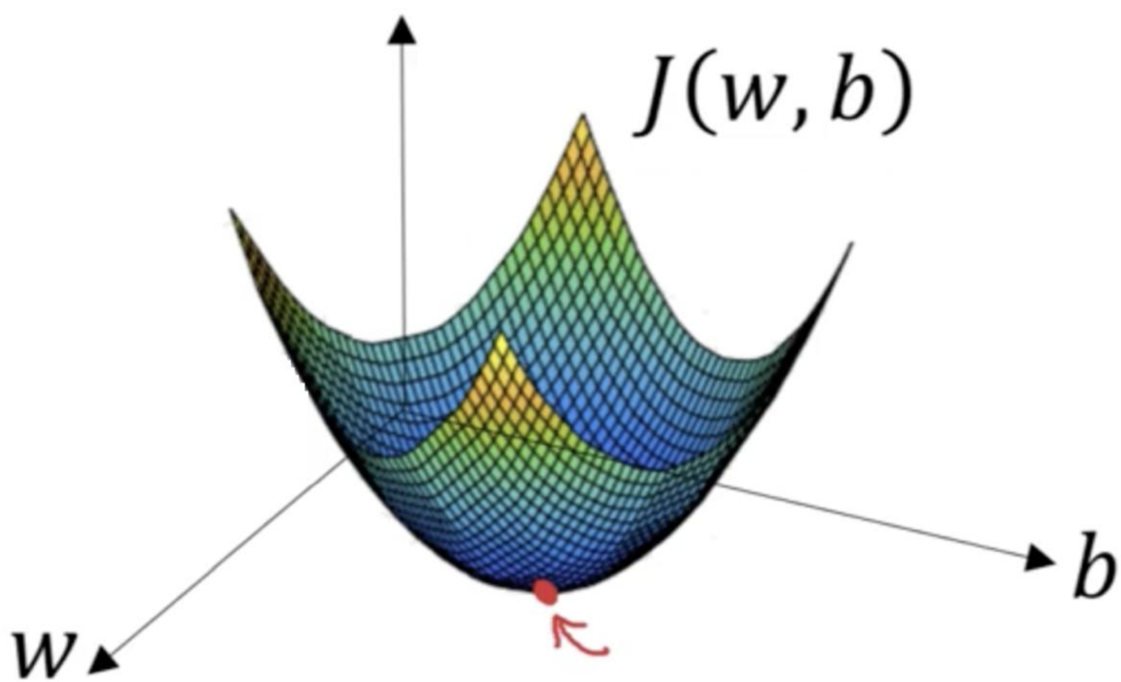
$$J(W,b) = \frac{1}{n_x} \sum_{i=1}^{n_x} L(y^{(i)}, \hat{y}^{(i)})$$

$$J(W,b) = \frac{1}{n_x} \sum_{i=1}^{n_x} \left(y^{(i)} - \hat{y}^{(i)} \right)^2$$

- Output
  - $\hat{y}^{(i)} \sim y^{(i)}$
  - W and b – Minimize J(W, b)

Gradient Descent

- Convex function
- Global optimum



Gradient Descent

- Forward pass
  - $\hat{y} = W^T X + b$
  - $L(y^{(i)}, \hat{y}^{(i)})$
  - $J(W, b)$
- Backward pass

$$dW = \frac{\partial J}{\partial W}$$

$$db = \frac{\partial J}{\partial b}$$

Gradient Descent

- Forward pass
  - $\hat{y} = W^T X + b$
  - $L(y^{(i)}, \hat{y}^{(i)})$
  - $J(W, b)$
- Backward pass

$$dW = \frac{\partial J}{\partial W}$$

$$db = \frac{\partial J}{\partial b}$$

$$W = W - \alpha * dW$$

$$b = b - \alpha * db$$

Gradient Descent

- Forward pass
  - $\hat{y} = W^T X + b$
  - $L(y^{(i)}, \hat{y}^{(i)})$
  - $J(W, b)$

- Backward pass

$$db = \frac{\partial J}{\partial b} = \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial b}$$

$$\frac{\partial J}{\partial \hat{y}} = \frac{2}{n_x} \times (y - \hat{y})$$

$$\frac{\partial \hat{y}}{\partial b} = 1$$

Gradient Descent

- Forward pass
  - $\hat{y} = W^T X + b$
  - $L(y^{(i)}, \hat{y}^{(i)})$
  - $J(W, b)$
- Backward pass

$$db = \frac{2}{n_x} \times (y - \hat{y}) \times (1)$$

$$db = \frac{2}{n_x} \times (y - \hat{y})$$

Gradient Descent

- Forward pass
  - $\hat{y} = W^T X + b$
  - $L(y^{(i)}, \hat{y}^{(i)})$
  - $J(W, b)$

- Backward pass

$$dW = \frac{\partial L}{\partial W} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial W}$$

$$\frac{\partial L}{\partial \hat{y}} = \frac{2}{n_x} \times (y - \hat{y})$$

$$\frac{\partial \hat{y}}{\partial W} = X$$

Gradient Descent

- Forward pass
  - $\hat{y} = W^T X + b$
  - $L(y^{(i)}, \hat{y}^{(i)})$
  - $J(W, b)$
- Backward pass

$$dW = \frac{2}{n_x} \times (y - \hat{y}) \times (X)$$

R^2

- Coefficient of determination
- Coefficient of multiple determination
- Strength of relationship
- Value between 0.0 – 1.0
- Percentage value
- Independent variable explains p percent of variation in dependent variable
- Independent variable reduces p percent of variation in dependent variable

Adjusted R^2

- R^2
  - Increase independent variables – Increase R^2
  - Increase independent variables – Constant R^2
- Adjusted R^2
  - Increase independent variables (then)
  - Increase model accuracy (then only)
  - Increase adjusted R2

Applications

- Inputs to response mapping
- Error reduction
- Prediction
  - House price prediction from observed dataset
- Forecasting
  - Weather forecasting from observed dataset

Regularization

- Lasso regression
- Ridge Regression
- Elastic Net Regression

Lasso Regression

- Least Absolute Shrinkage Selector Operator
- L1 regularization technique
- Reduce coefficients
- Feature selection
  - Select important features
  - Reduce coefficients of others to zero
- Suitable for more number of features

Ridge Regression

- L2 regularization technique
- Reduce coefficients
- Reduce model complexity
- Prevent multicollinearity

Elastic Net Regression

- L1 and L2 regularization technique

Advantages

- Good for linearly separable data
- Easier to implement and interpret
- Efficient to train
- Handle over-fitting
  - Dimensionality reduction techniques
  - Cross validation
  - Regularization
- Extrapolation of dataset

Disadvantages

- Assumption of linearity
  - Independent variables
  - Dependent variables
- Prone to noise
- Sensitive to outliers
- Prone to multicollinearity

Multicollinearity

- Correlated independent variables
- Example
  - Independent variables
    - Radius of a circle
    - Circumference of a circle
  - Radius and circumference – Correlated

Questions?

Thank you