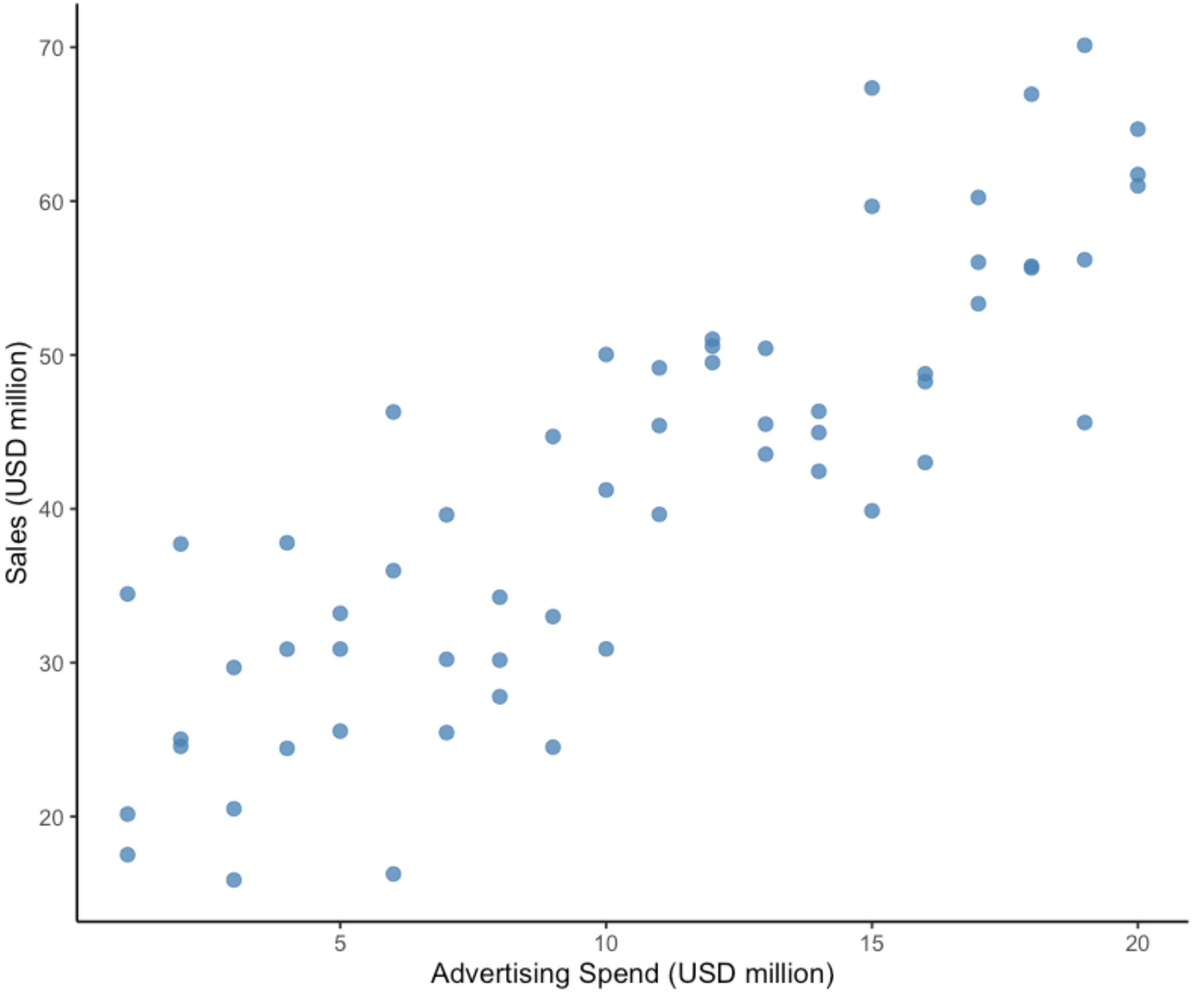
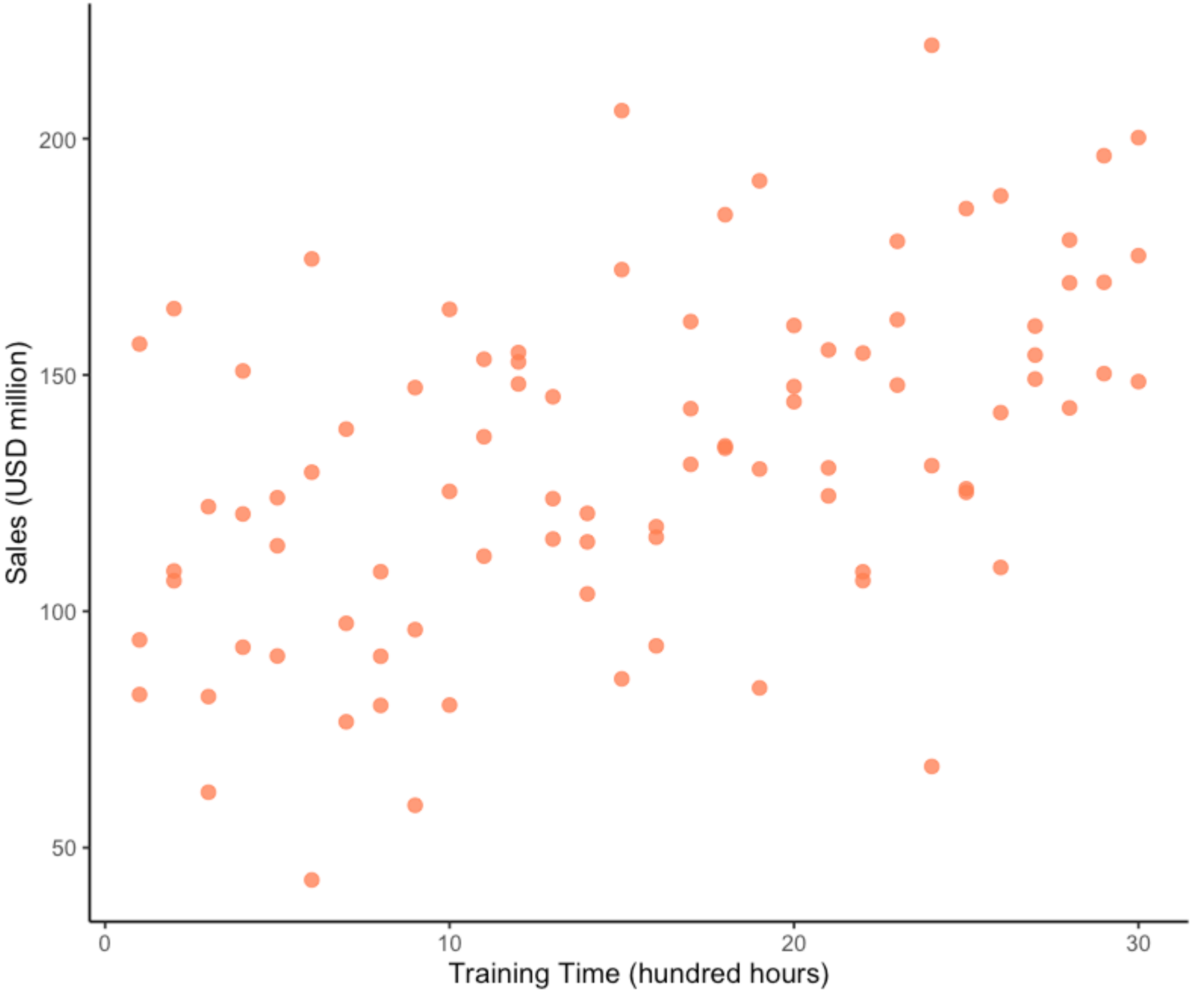
Regression
Part A

Ulink MM Club

Alan, Frank

Introduction





Linear Regression

General formula

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

where β_0 denotes the intercept, β_1 denotes the gradient, and ϵ_i denotes the error term (or residue).

Key Assumptions

- Observations are independent of each other.
- The residue is normally distributed.

Likelihood Functions

A likelihood function measures how well a statistical model explains observed data by calculating the probability of seeing that data under different parameter values of the model, written as:

$$\mathcal{L}(\theta|x)$$

which denotes the probability of obtaining parameter θ given observation x.

Maximum Likelihood Estimation

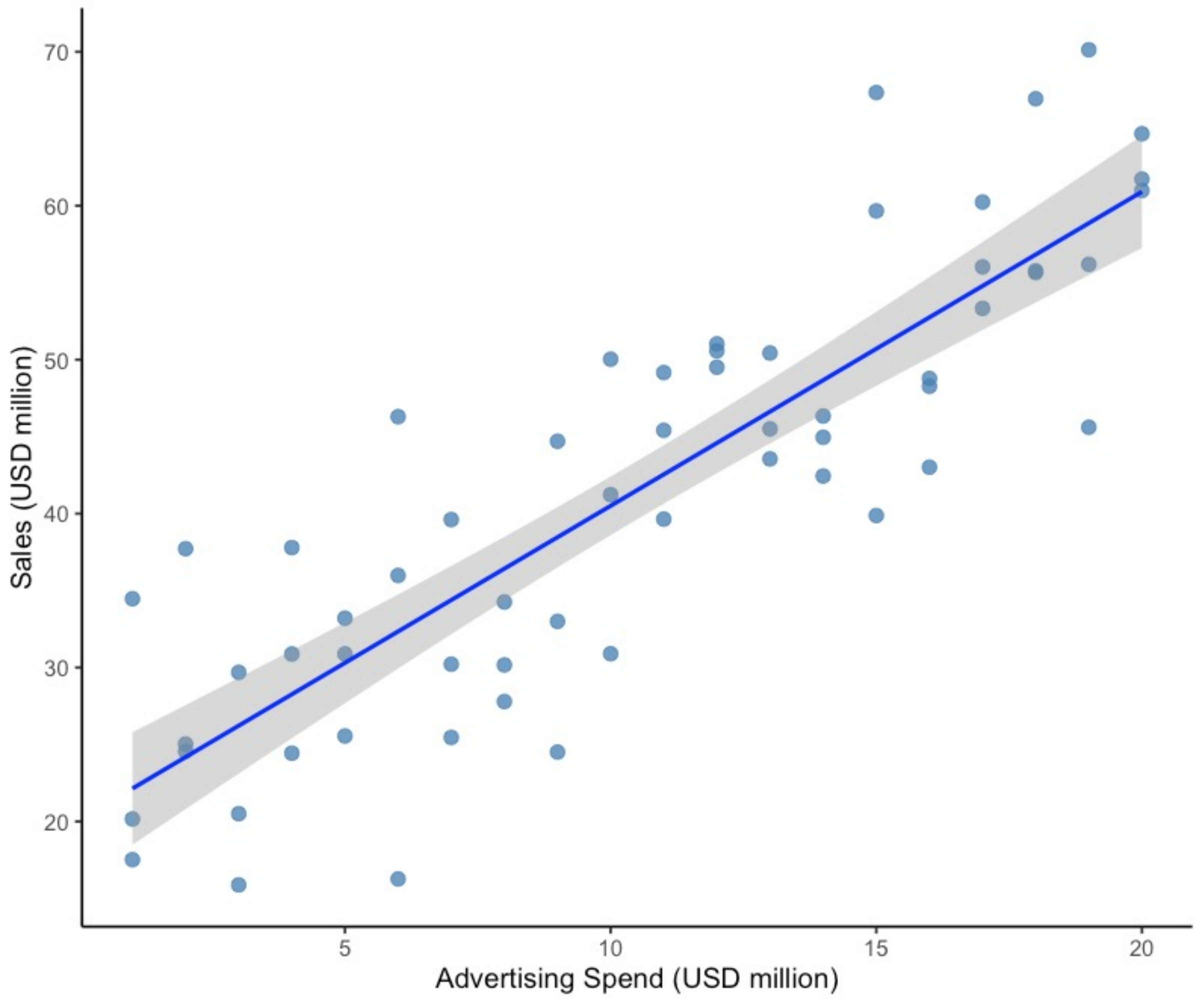
We aim to maximise this $\mathcal{L}(\theta|x)$, so that we have high probability of getting θ given x, that is to say, the parameter θ best fits the observations.

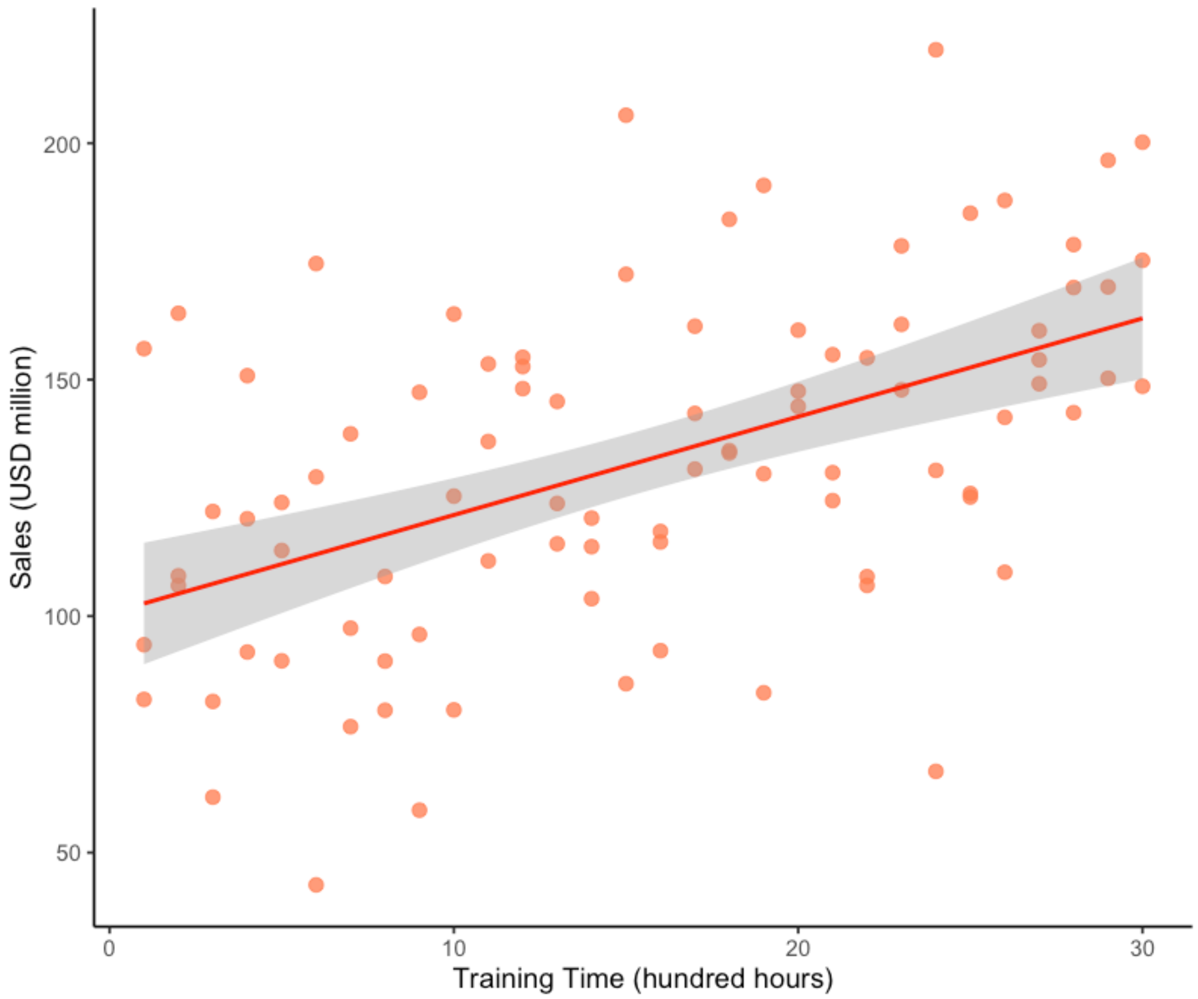
Conclusion for Linear Regressions

Maximum likelihood estimation is identical to least square estimation.

Mean squared error (MSE): $MSE(\beta_0, \beta_1) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$.

Results


70
60
50
40
30
20
Sales (USD million)
5
10
15
20
Advertising Spend (USD million)


200
150
100
50
Sales (USD million)
0
10
20
30
Training Time (hundred hours)

Evaluating the Models

Coefficient of determination $\boldsymbol{R^2}$

$$SS_{res} = \sum_{i=1}^{n} (y_i - \widehat{y_i})^2 \,, SS_{tot} = \sum_{i=1}^{n} (y_i - \overline{y_i})^2$$

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

If the model exactly fits the data, $SS_{res} = 0$ and $R^2 = 1$,

If the model always estimates output to be $\overline{y}$, $SS_{res} = SS_{tot}$ and $R^2 = 0$.

Evaluating the Models

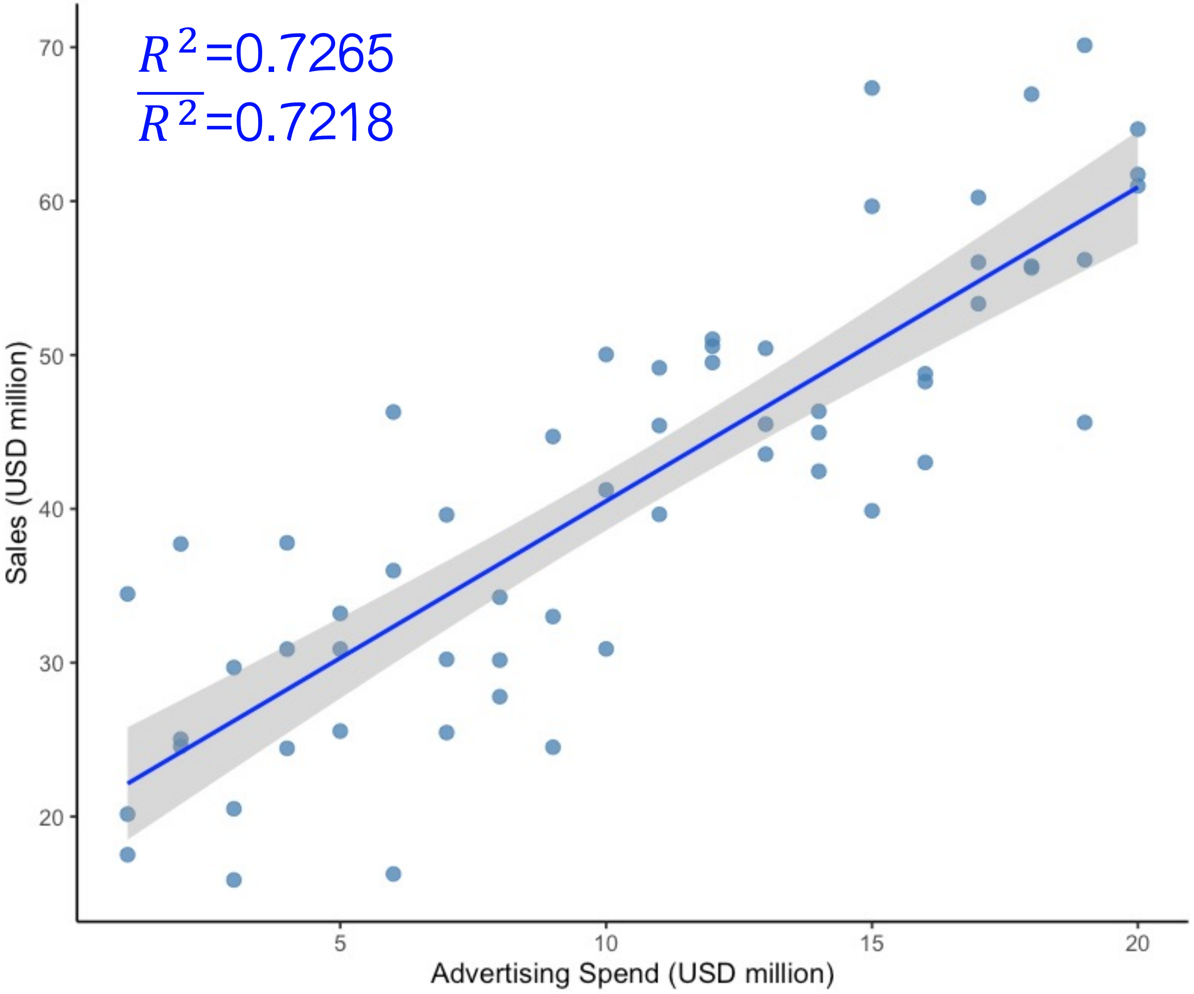
Adjusted $\boldsymbol{R^2}$

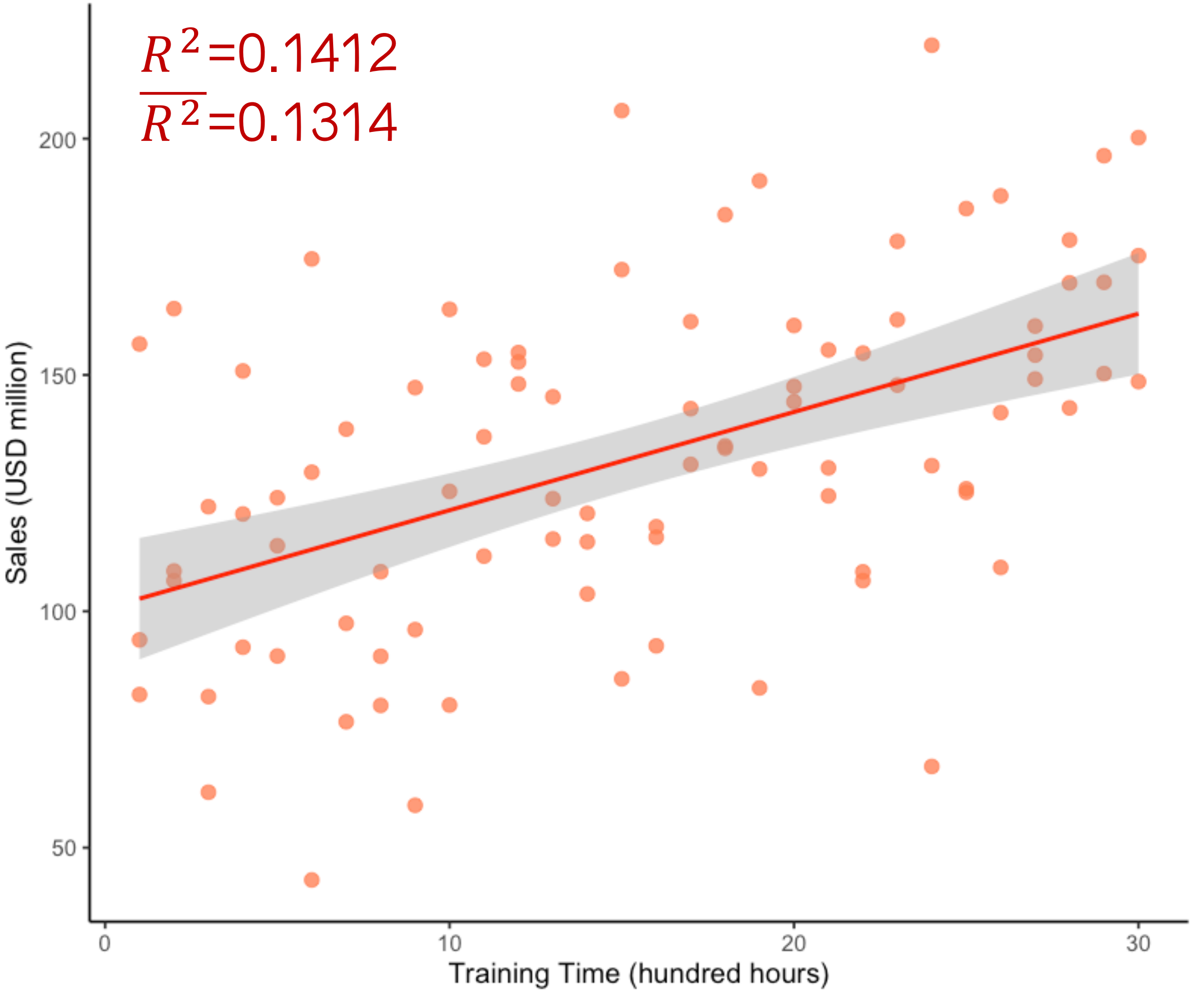
R^2 measures how well the model fits the historical data, but not how well the model will forecast future data. In addition, adding any variable tends to increase the value of R^2, even if that variable is irrelevant.

$$\overline{R^2} = 1 - (1 - R^2)\frac{T - 1}{T - k - 1}$$

where T denotes the number of observations and k denotes the number of predictors (independent variables).

Results


R^2=0.7265
$\overline{R^2}$=0.7218
70
60
50
40
30
20
Sales (USD million)
5
10
15
20
Advertising Spend (USD million)


R^2=0.1412
$\overline{R^2}$=0.1314
200
150
100
50
Sales (USD million)
0
10
20
30
Training Time (hundred hours)

Strengths and Weaknesses

Strengths:

- The model gives a direct relationship between predictors and the target (e.g. each 1 unit increase in X adds 2.5 units to Y);
- Fast to train and test, even on large datasets, as it has a closed-form solution;

Weaknesses:

- Sensitive to Outliers;
- When the number of predictors is large relative to observations, it can overfit unless regularization (e.g., Ridge, Lasso) is applied.


Thank you