

Assignment Instructions

Consider a simple 2D dataset where the data is easy to visualize as a scatter plot and models are easy to visualize as a line that attempts to fit the data points.

The data points are:

(X, Y): age (months) & Height (inches) of a child

(0, 19.75), (1, 21.4), (2, 23.5), (3, 23.1), (4, 23.8), (5, 25), (6, 26.4), (7, 27.25)

Obtain the Scatter Plot of the Dataset.

Using the formula $\mathbf{w} = \mathbf{X}^\dagger \mathbf{y}$ where $\mathbf{X}^\dagger = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$, find the parameters of a linear model, i.e. (a line) \mathbf{w} .

Note that each data point

$$\mathbf{x}_i = (x_0, x_1) = (1, x_i), i = 1, \dots, 8$$

Plot the line obtained together with the data points showing a reasonable fit to the data.

Fitting a Line to a Scatter Plot from Given Data

Table of Contents

| | |
|---|---|
| Data Provided | 1 |
| Plotting the Data: Age and Height of Child | 1 |
| Obtaining Weights for the Best Fit Line Using Linear Regression | 2 |
| Plotting the Data and the Best Fit Line | 2 |

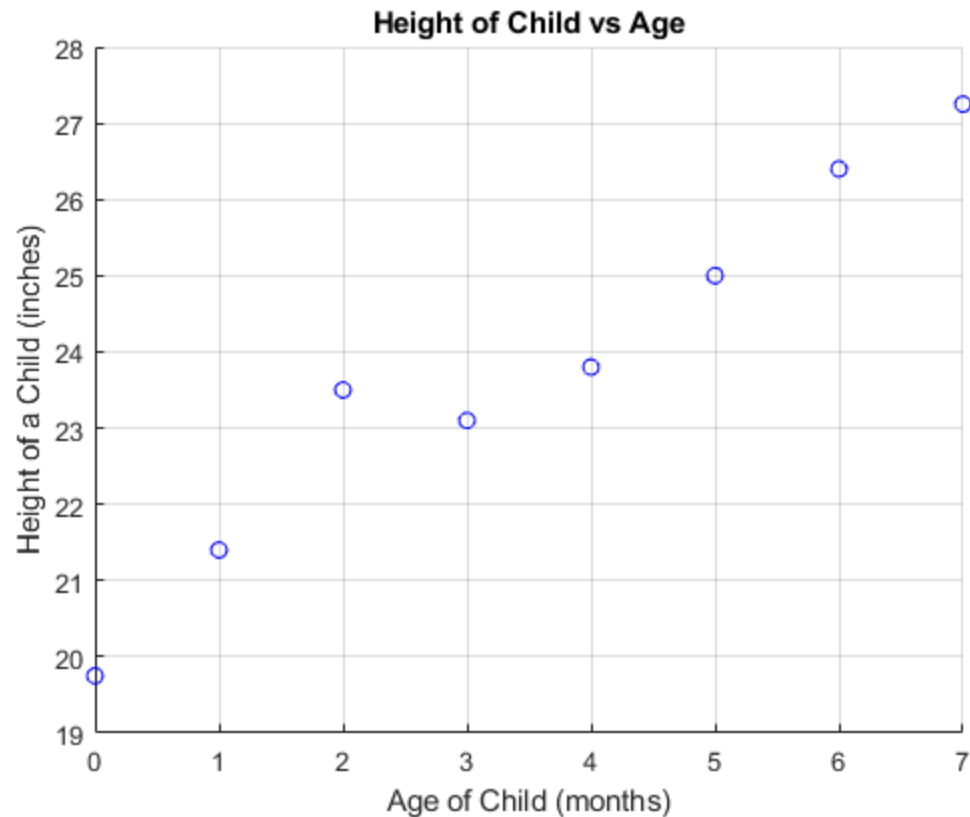
Data Provided

Below is the data given for this assignment. x vector below shows the age of child in months. y vector below shows the child's height in inches

```
x = [0; 1; 2; 3; 4; 5; 6; 7];  
y = [19.75; 21.4; 23.5; 23.1; 23.8; 25; 26.4; 27.25];
```

Plotting the Data: Age and Height of Child

```
figure(1)  
scatter(x, y, 'b');  
grid on  
%hold on  
title('Height of Child vs Age');  
xlabel('Age of Child (months)');  
ylabel('Height of a Child (inches)');
```



Obtaining Weights for the Best Fit Line Using Linear Regression

```
X_fit = [1 0; 1 1; 1 2; 1 3; 1 4; 1 5; 1 6; 1 7]; % X values for
linear
% fit plot line

X_Tpose = transpose(X_fit); % transpose of X_fit vector
X_cross = ((X_Tpose*X_fit)^-1)*X_Tpose; % inverse of X_fit
w = X_cross*y; % w parameters to be used on function fit line
disp('The weights given by the algorithm to produce a best fit line
for the data are below');
disp(w);
```

The weights given by the algorithm to produce a best fit line for the data are below

*20.3292
0.9845*

Plotting the Data and the Best Fit Line

```
figure(2)
scatter(x, y, 'b');
```

```
grid on
hold on
title('Height of Child vs Age');
xlabel('Age of Child (months)');
ylabel('Height of a Child (inches)');

fprintf('y-intercetp is %g and slope is %g\n',round(w(1),4),round(w(2),4));
fprintf('y = %gx + %g\n',round(w(2),4),round(w(1),4));
% creating a fit line to plot againts the data plot to show this is
% the best fit line.
% the weights, (w's), used to produced this plot are the w's generated
% by the algorithm above.
x_fit_plot = (0:7);
y_fit = w(2,1)*x_fit_plot + w(1,1); % funciton line used for fit
plot(x_fit_plot, y_fit, 'm')
legend('height vs age', 'line fit function', 'location','northwest')
text(3,21,'best fit line','FontSize',12);
txt = ['y = 0.9845x + 20.3292'];
text(3,20.5,txt,'FontSize',12);
hold off
```

y-intercetp is 20.3292 and slope is 0.9845

$y = 0.9845x + 20.3292$

