Dimensions and Hyperplanes

Visualizing n −dimensions

- 1D, 2D and 3D being!

- **How transforming data from low-dimensional space to high-dimensional space is helpful?**
  - Polynomial features
  - Polynomial degree

- We can fit a linear line on our data to find the optimal **decision boundary.**

Fitting a Line to the Data

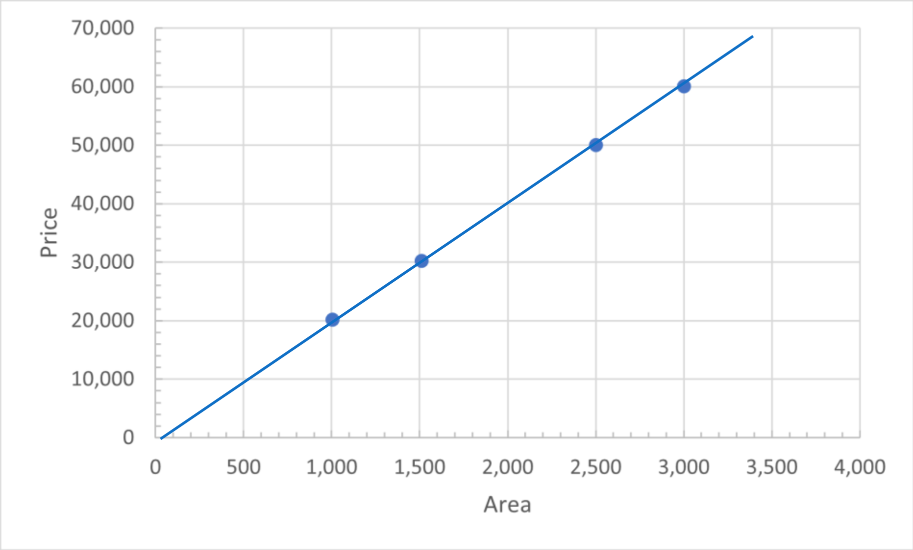
- Suppose we have the following training data

$Area\ (m^2)$	$Price\ (\$)$
1,510	30,250
1,005	20,150
2,500	50,050
3000	60,050

- We need to predict values for the following test data

$Area\ (m^2)$	$Price\ (\$)$
500	?
2,000	?
3,500	?

Predictions?



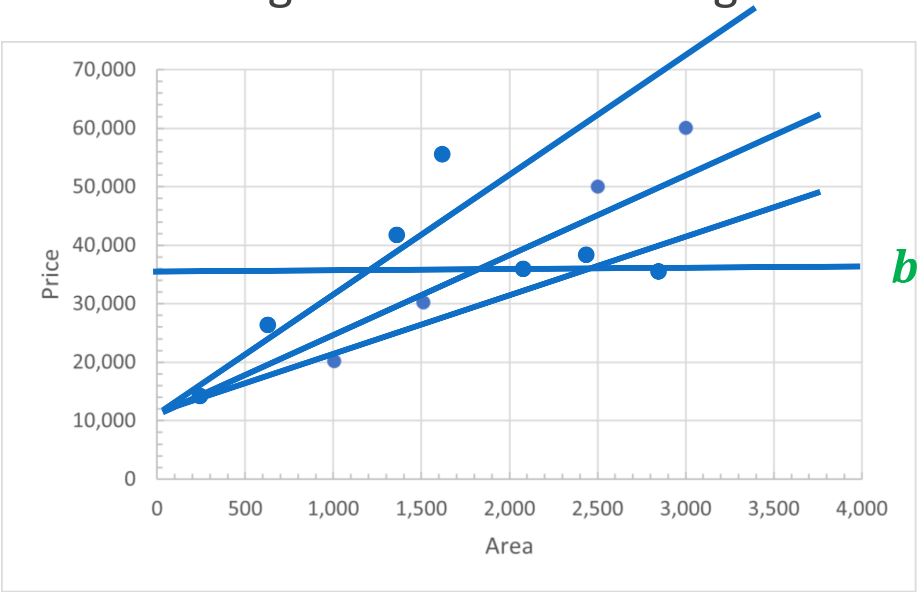
What kind of relationship exists between the X (Area) and y (Price)?

We don't need any fancy algorithm to predict the price for test data.

Many real-world data has a linear or "somewhat linear" relationship between input and output.

Fitting a Line to the Data

- What if training data was something like:



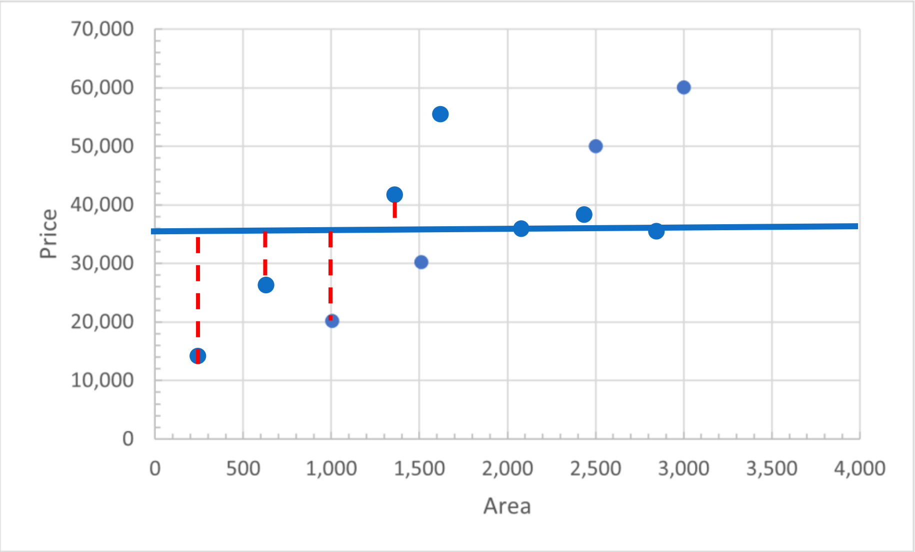
Starting with horizontal line is good idea that cuts through average value of y

Let's call this average value b

Where should we draw our "decision line"?

Fitting a Line to the Data

- What if training data was something like:



$\hat{y} = \ b = 3500$

$(\hat{y}_1 - y_1) + (\hat{y}_2 - y_2) + (\hat{y}_3 - y_3) + (\hat{y}_4 - y_4)$

How well this line "fits" the data?

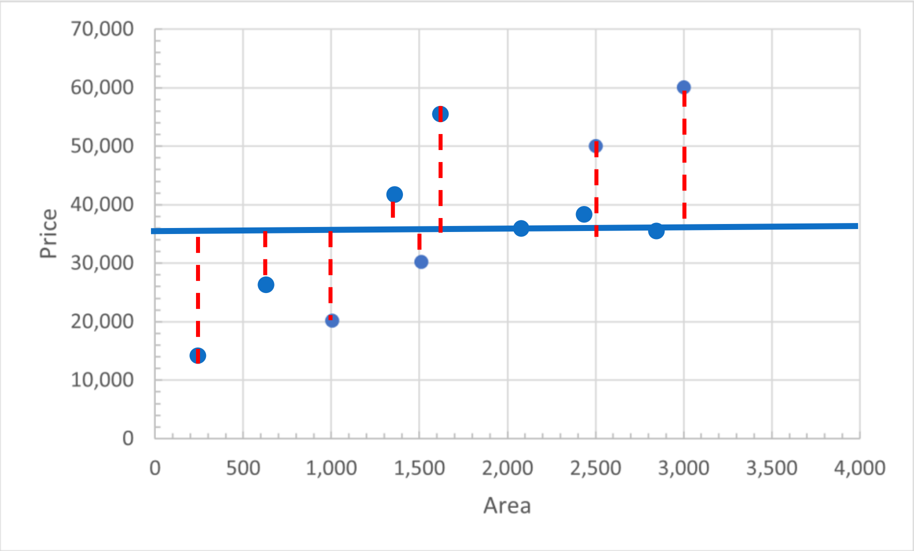
In other words, if you were to use this "decision line" to predict the y value for a given x value, how good the prediction would be?

$\hat{y}$ is smaller than y. The outcome would be negative, which will subtract from the total making the overall fit appear better than it really is!

Take difference of actual y value from predicted $\hat{y}$ value

Fitting a Line to the Data

- What if training data was something like:



$$\hat{y} = \; b = 3500$$

$$(\hat{y}_1 - y_1)^2 + (\hat{y}_2 - y_2)^2 + (\hat{y}_3 - y_3)^2 + (\hat{y}_4 - y_4)^2 + \ldots + (\hat{y}_n - y_n)^2$$

We square each term to make it non-negative

This is known as "Sum of Squared Residuals" or "Sum of Squared Errors" or "Loss" or "Cost".

$$SSE = \sum_{i=1}^{n} (h(x_i) - y_i)^2$$

Take Aways

- **How much error would there be between actual and predicted values?**
  - **Its how far away the line is from the observed values.**

- **How to compute error numerically?**

$$\boldsymbol{SSE = \sum_{i=1}^{n} (h(x_i) - y_i)^2}$$

Squared term always gives the positive value.

$$\boldsymbol{SAE = \sum_{i=1}^{n} |h(x_i) - y_i|}$$

Absolute value is always positive.

Take Aways

- **But SSE and SAE is not always quantifying the real picture…**
  - **Two datasets with different number of data points can have same SSE or SAE.**

- **Fix: Take Mean**

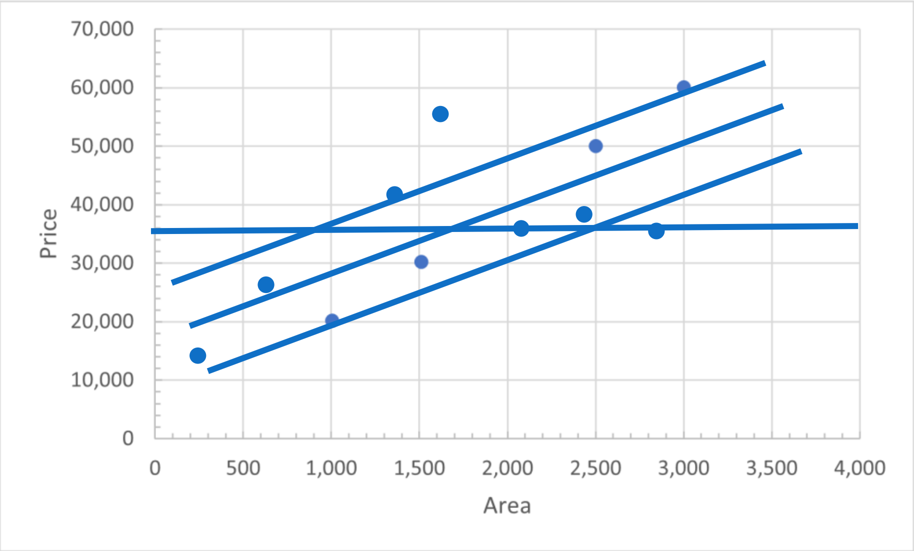
$$MSE = \frac{1}{n}\sum_{i=1}^{n}(h(x_i) - y_i)^2$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|h(x_i) - y_i|$$

Do we want to maximize or minimize this cost function?

Minimize

But We can Also Rotate the Line!



$$y = m \times x + b$$

The Slope

y intercept

Which line is better fit to the data?

How do we find sweet spot between slope and y-intercept?

Start with Generic Line Equation

We want to find the optimal values for m and b which minimizes the sum of squared residuals (or we say, minimize our cost function which is MSE)

Where the line would be drawn if $b = 0$?

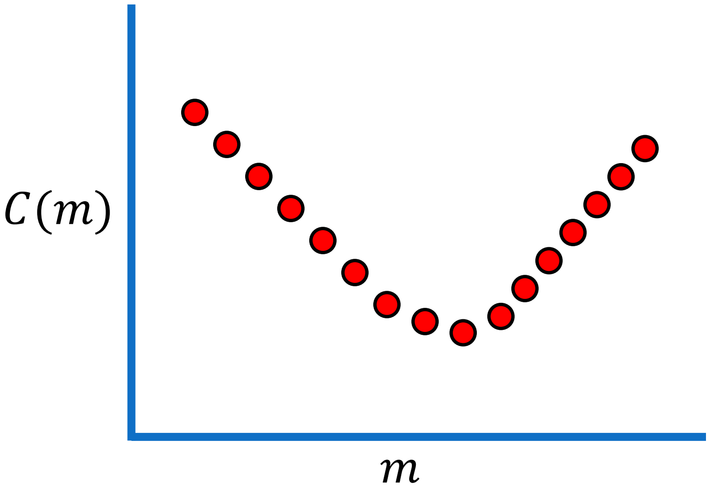
It will start from the origin, but its slope may vary depending on the value of m.

Let's Plot the Cost Function

$$C(\,) = \frac{1}{n}\sum_{i=1}^{n}((m \times x_i + b) - y_i)^2$$

To minimize this cost, we must either change the value of m or the value of b.

We can minimize the cost function by setting up $b = 0$ and then optimizing the value of m (i.e., optimizing both parameters independently).



$$C(m) = \frac{1}{n}\sum_{i=1}^{n}((m \times x_i) - y_i)^2$$

This is a convex function, and we just find in which direction we must move and how bigger the step should be!

Optimization problem!
Many algorithms to find this minima (Such as Gradient Descent)

Demo

- https://www.desmos.com/calculator/m1skuqsxkj

Regression Analysis

Regression Analysis

- Is used to:
  - Predict the value of a ***dependent variable (aka output)*** based on the value of ***at least one independent variable (aka input)***
  - Explain the ***impact of changes in an independent variable*** on the ***dependent variable***

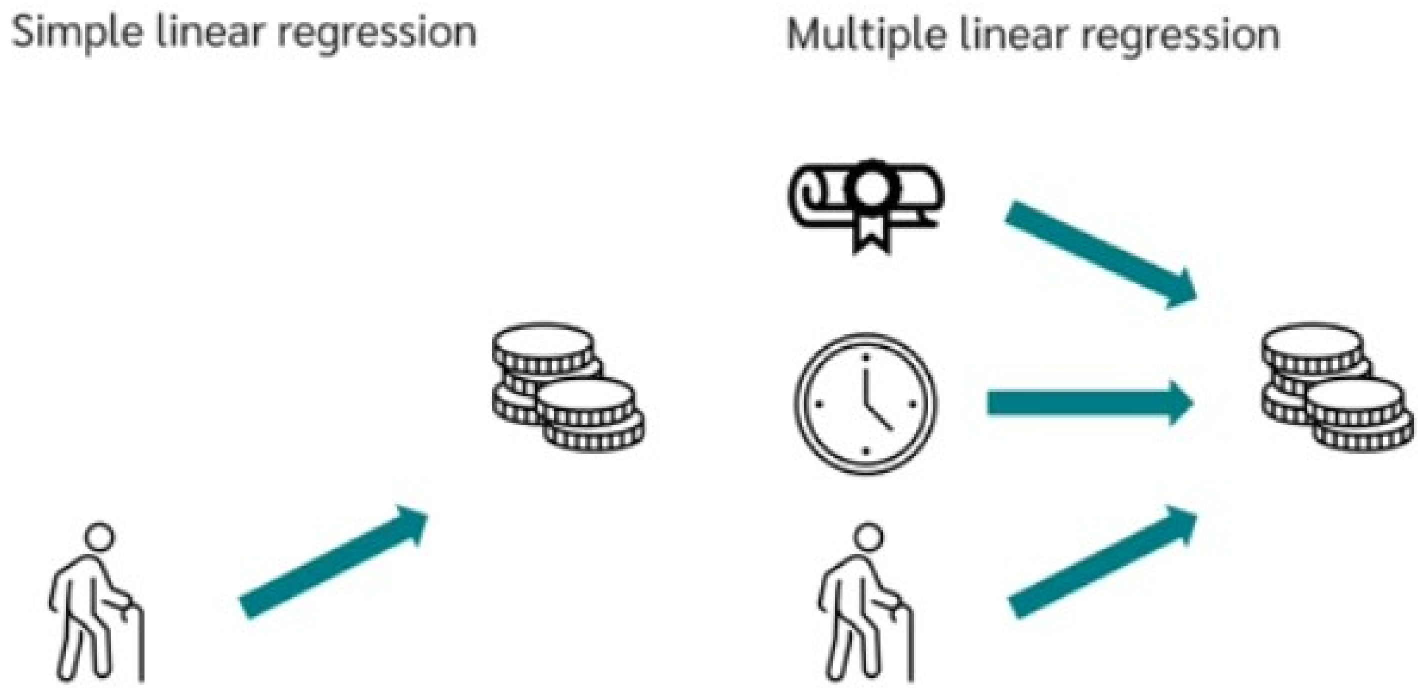
- ***Dependent Variable:***
  - The variable we wish to explain
- ***Independent Variable:***
  - The variable used to explain the dependent variable

Regression Analysis

- The Key Steps for Regression:
  1. List all the variables available for making the model
  2. Establish a dependent variable of interest
  3. Examine (if possible), visual relationships between variables of interest
  4. Find a way to predict dependent variable using the other variables

- The regression model is described as a linear equation, that follows:
  - ***y*** is the dependent variable (the variable being predicted)
  - ***X*** is the independent variable (or the predictor variable)
  - There could only be one dependent variable (y) in the regression equation.

Simple Linear Regression vs Multiple Linear Regression



Linear Regression

ONE INDEPENDENT VARIABLE

Linear Regression with one independent variable

- Also called ***"Simple Regression"*** or ***"Simple Linear Regression"*** or ***"Ordinary Least Squares (OLS)"***
- Establishes the ***linear relationship*** between two variables based on a line of best fit.
- Its graphically depicted using a ***straight line with a slope***, defining how the change in one variable impacts the change in the other
- The ***y-intercept*** of a linear regression represents the value of ***one variable when the value of the other is zero (Recall bias)***

Example 1

ORDINARY LEAST SQUARES: METHOD 1

Predict Glucose Level, Given Age

- **Remember:** In regression, output is normally a continuous value (as opposed to classification, where the output is a discrete/categorical value)

$Age\ (X)$	$Glucose\ Level\ (y)$
43	99
21	65
25	79
42	75
57	87
59	81

$Age\ (X)$	$Glucose\ Level\ (y)$
55	?

Many forms of the equation:

$\hat{y} = b_0 + b_1 X_1$ | $\hat{y} = mx + b$ | $\hat{y} = \theta_0 + \theta_1 X_1$ | $\hat{y} = w_1 X_1 + b$

Keeping in view the above equation, we already know the value of X (55), but we do not know the values of *b* and *m*. We can find the values of these two parameters by using *known values of x and y*.

$$b = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

$$m = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

Recall *Pearson's Correlation Coefficient*

Predict Glucose Level, Given Age

$$\hat{y} = mx + b$$

	$Age\ (X)$	$Glucose\ Level\ (y)$	Xy	X^2
	43	99	4257	1849
	21	65	1365	441
	25	79	1975	625
	42	75	3150	1764
	57	87	4959	3249
	59	81	4779	3481
Σ	**247**	**486**	**20485**	**11409**

$Age\ (X)$	$Glucose\ Level\ (y)$
55	?

$$b = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

$$m = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$b = \frac{(486)(11409) - (247)(20485)}{6(11409) - (247)^2}$$

$$b = \frac{4848979}{7445} = 65.14$$

$$m = \frac{6(20485) - (247)(486)}{6(11409) - (247)^2}$$

$$m = \frac{2868}{7445} = 0.385335$$

$$\hat{y} = 0.385225x + 65.14$$

Model

Predict Glucose Level, Given Age

$$\hat{y} = mx + b$$

$$\hat{y} = 0.385225x + 65.14$$

Model

$Age\ (X)$	$Glucose\ Level\ (y)$	Xy	X^2
43	99	4257	1849
21	65	1365	441
25	79	1975	625
42	75	3150	1764
57	87	4959	3249
59	81	4779	3481
$\sum$ **247**	**486**	**20485**	**11409**

Alternate method to calculate b

$$b = \frac{\sum y - m(\sum x)}{n}$$

$$b = \frac{486 - 0.385225(247)}{6}$$

$$b = \frac{486 - 95.150575}{6} = 65.14$$

$Age\ (X)$	$Glucose\ Level\ (y)$
55	?

$$\hat{y} = 65.14 + (0.385225 \times 55)$$

$$\hat{y} = 86.327$$

The glucose level for the given age 55 is 86.327

Points to note…

- The method that we have used is called ***Least Squares Method*** *(aka Line of Best Fit Equation)*
- There are other methods as well (such as ***Neural Network***, or ***Mean and Covariance***)

Why different equation forms?

- **In Statistics:**

$$\hat{y} = b_0 + b_1 X_1$$

In simple words, this is called "Regression coefficient". It tells how much impact of feature X_1 has on the output.

- **In Linear Algebra:**

$$\hat{y} = mx + b$$

- $Slope = m = b_1$
- $y - intercept = b = b_0$

What if we have more than one input/independent variables?

$$\hat{y} = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_k X_k$$

OR

$$\hat{y} = w_0 x_0 + w_1 x_1 + \cdots + w_n x_n$$

Remember this form of equation…

Example 2

METHOD 2: VARIANCE AND COVARIANCE

Covariance, Variance Recap…

- The estimated parameters of independent variables will be related to the covariance of the dependent variable.
- The ***covariance*** of a variable ***with itself*** simplifies to just ***variance***.
- The ***variance-covariance matrix*** (all the pairwise covariances and all the variances) is enough information to compute all the OLS coefficients, ***except the intercept***, which requires information about the ***means***.

Simple Linear Regression: Solving through Covariance and Variance

$$\hat{y} = mx + b$$

x	y
1	1
2	2
3	1.3
4	3.75
5	2.25

Step 1: Compute Means and Covariance

$$\bar{x} = \frac{1}{n}\sum x_i \qquad \bar{x} = \frac{1}{5}(1 + 2 + 3 + 4 + 5) = 3$$

$$\bar{y} = \frac{1}{n}\sum y_i \qquad \bar{y} = \frac{1}{5}(1 + 2 + 1.3 + 3.75 + 2.25) = 2.06$$

$$cov(x, y) = \frac{1}{n-1}\sum(x_i - \bar{x})(y_i - \bar{y}) \qquad cov(x, y) = \frac{1}{4}[(1-3)(1-2.06) + \cdots + (5-3)(2.25-2.06)]$$

$$cov(x, y) = 1.0625$$

$$var(x) = \frac{1}{n-1}\sum(x_i - \bar{x})^2 \qquad var(x) = \frac{1}{4}[(1-3)^2 + \cdots + (5-3)^2$$

$$var(x) = 2.5$$

$$m = \frac{cov(x, y)}{var(x)} \qquad m = \frac{1.0625}{2.5} = 0.425 \qquad b = \bar{y} - m\bar{x} \qquad b = 2.06 - 0.425 \times 3 = 0.785 \qquad \hat{y} = 0.425x + 0.785$$

Simple Linear Regression: Solving through Covariance and Variance

x	y
1	1
2	2
3	1.3
4	3.75
5	2.25

Step 2: Save this model and use it to predict new values for a given x

$$\boxed{\hat{y} = \mathbf{0.425}x + \mathbf{0.785}}$$

You can also use correlation and standard deviation:

$$\boxed{m = r\frac{S_y}{S_x}} \quad \boxed{b = \bar{y} - b\bar{x}}$$

But we are not going in much details of that method…

x	y
6	?

$$\boxed{y = \mathbf{0.425} \times \mathbf{6} + \mathbf{0.785} = \mathbf{3.335}}$$

Evaluation with R^2 Score

- $R^2 = 1 - \frac{u}{v}$
  - Where u is residual sum of squares $(\sum_i^n (y_i - \hat{y}_i)^2)$
  - v is total sum of squares$(\sum_i^n (y_i - \bar{y})^2)$

- Best value for R^2 can be 1.

Useful Resources

- https://www.investopedia.com/terms/r/regression.asp

Assignment 2: Task 1

- Use your images dataset to train a linear regression model.
- Use age as label.
- Compute R^2 on test split.
- Compute MSE or test split.

Book Reading

- Murphy – Chapter 1.4