Multivariate Regression

Lesson

AI Academy

Multivariate Regression

- What if we have several inputs?
  - E.g. predict Google's price using prices for Amazon, Microsoft and eBay.
- This becomes a multivariate linear regression problem
- We model that as: $y = w_0 + w_1 x_1 + \cdots + w_k x_k + \varepsilon$

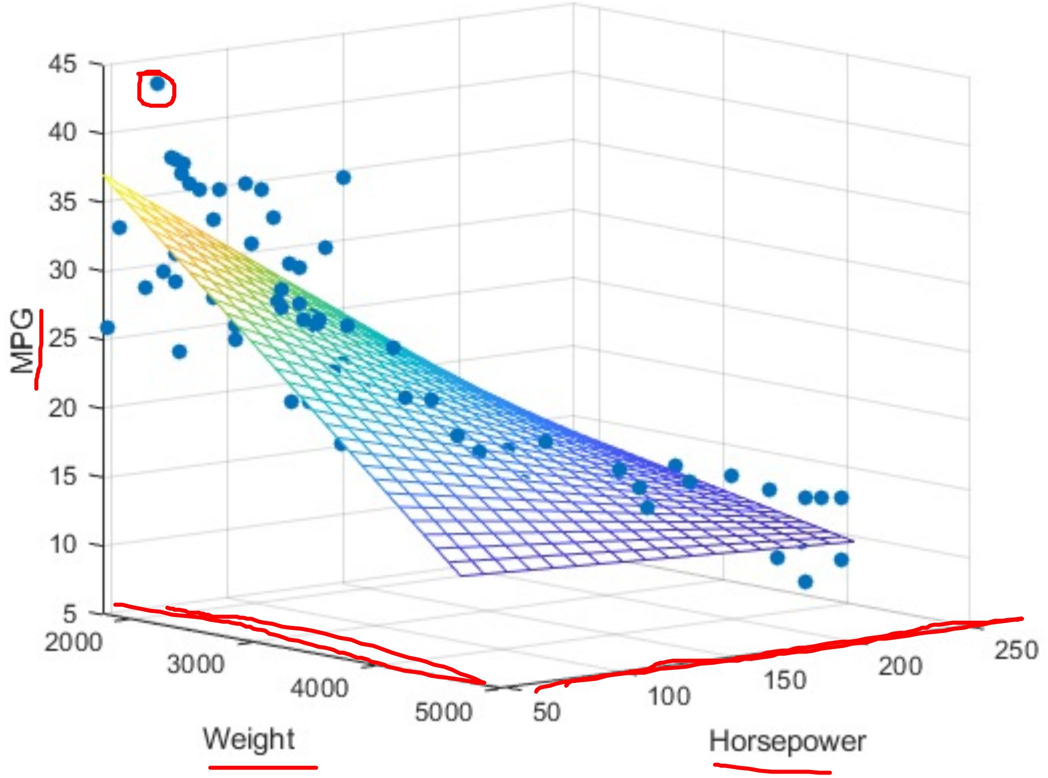
Google's Stock Price

Amazon's Stock Price

Microsoft's Stock Price

Multivariate Regression: Conceptual Overview

- We can map a single y variable to multiple x features.
- This allows for interpretation of feature weights.



https://medium.com/analytics-vidhya/new-aspects-to-consider-while-moving-from-simple-linear-regression-to-multiple-linear-regression-dad06b3449ff

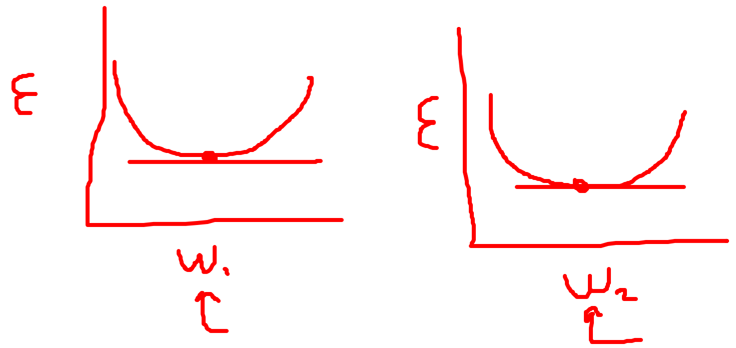
Interpretability

- Linear Regression is *interpretable*:
  - The model parameters have meaning
- E.g. Predicting Sepal Length from other Iris attributes

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.19058    0.09705  43.181  < 2e-16 ***
## Petal.Length  0.54178    0.06928   7.820 9.41e-13 ***
## Petal.Width  -0.31955    0.16045  -1.992   0.0483 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.g Petal Length positively predicts Sepal Length

Multivariate Regression



- What if we have several inputs?
  - E.g. predict Google's price using prices for Amazon, Microsoft and eBay.
- This becomes a multivariate linear regression problem
- We model that as: $y = w_0 + w_1 x_1 + \cdots + w_k x_k + \varepsilon$
- To solve for all w_i, take the partial derivative of error w.r.t each weight and set it equal to 0.
- Solve the resulting system of equations.

Example: Systems of Equations

x1	y
-1	1
0	-1
2	1

$$y = w_0 + w_1 x_1 + \epsilon$$

SSE

$$J(\mathbf{w}) = \sum_i (y_i - (w_0 + w_1 x_i))^2$$

$$\frac{\delta}{\delta w_0} J(\mathbf{w}) = \sum_i 2(y_i - (w_0 + w_1 x_i)) = 0$$

Example: Systems of Equations

	x1	y
1	-1	1
2	0	-1
3	2	1

$$y = w_0 + w_1 x_1 + \epsilon$$

$$J(\mathbf{w}) = \sum_i (y_i - (w_0 + w_1 x_i))^2$$

$$\frac{\delta}{\delta w_0} J(\mathbf{w}) = \sum_i 2(y_i - (w_0 + w_1 x_i)) = 0$$

$$2(y_1 - (w_0 + w_1 x_1)) + 2(y_2 - (w_0 + w_1 x_2)) + 2(y_3 - (w_0 + w_1 x_3)) = 0$$

$$2(1 - (w_0 = w_1)) + 2(-1 - (w_0 + w_1(0))) + 2(1 - (w_0 + w_1(2))) = 0$$

$$\boxed{1 - 3w_0 - w_1 + = 0}$$

Example: Systems of Equations

x1	y
-1	1
0	-1
2	1

$$y = w_0 + w_1 x_1 + \epsilon$$

$$J(\mathbf{w}) = \sum_i (y_i - (w_0 + w_1 x_i))^2$$

$$\frac{\delta}{\delta w_0} J(\mathbf{w}) = \sum_i 2(y_i - (w_0 + w_1 x_i)) = 0$$

Repeat with d/dw_1… and solve the two equations

Linear Algebra Solution

Is there a more efficient way to solve for **w**?

Answer: We can encode the set of linear equations using matrices

Standard Linear Regression:

$$\boldsymbol{w} = \begin{bmatrix} w_1 \\ w_2 \\ \cdots \\ w_k \end{bmatrix} \qquad \boldsymbol{x_i} = \begin{bmatrix} x_{i_1} \\ x_{i_2} \\ \cdots \\ x_{i_k} \end{bmatrix} \qquad y_i = \boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x_i} = w_1 x_{i_1} + w_2 x_{i_2} + \cdots + w_k x_{i_k}$$

Linear Algebra Solution

Is there a more efficient way to solve for **w**?

Answer: We can encode the set of linear equations using matrices

Standard Linear Regression:

$$\nabla_w\left(J(\boldsymbol{w})\right) = -2\sum_{i=1}^{n}\left(y_i - \boldsymbol{w}^T\boldsymbol{x_i}\right)\boldsymbol{x_i} = \boldsymbol{0}$$

$$\nabla\boldsymbol{w}\left(J(\boldsymbol{w})\right) = \begin{bmatrix} \frac{d}{dw_1}J(\boldsymbol{w}) \\ \frac{d}{dw_2}J(\boldsymbol{w}) \\ \cdots \\ \frac{d}{dw_k}J(\boldsymbol{w}) \end{bmatrix}$$

$$\boldsymbol{w} = \begin{bmatrix} w_1 \\ w_2 \\ \cdots \\ w_k \end{bmatrix} \qquad \boldsymbol{x_i} = \begin{bmatrix} x_{i_1} \\ x_{i_2} \\ \cdots \\ x_{i_k} \end{bmatrix}$$

$$y_i = \boldsymbol{w}^T\boldsymbol{x_i} = w_1x_{i_1} + w_2x_{i_2} + \cdots + w_kx_{i_k}$$

Linear Algebra Solution

Is there a more efficient way to solve for **w**?

Answer: We can encode the set of linear equations using matrices

Standard Linear Regression:

$$\nabla_w\big(J(\mathbf{w})\big) = -2\sum_{i=1}^{n}\left(y_i - \mathbf{w}^T \mathbf{x}_i\right)\mathbf{x}_i = \mathbf{0}$$

Solution: $w = (X^T X)^{-1} X^T y$

Where **X** is an *n* by *d* matrix with rows corresponding to *n* examples and *d* columns to inputs.

$$X = \begin{array}{c} \\ x_1 \\ x_2 \\ \\ \end{array} \begin{array}{cc} f_1 & f_2 \\ \end{array} \left[\begin{array}{cc} 0 & \\ & \\ & \\ \end{array}\right]$$