Multivariate Linear Regression

Dave Miller

2/13/2019

Looking at a new dataset: Ames housing

Ames housing data contains the sale price and 80 variables of house information for 2930 properties sold in Ames, IA between 2006-2010:

```
Ames <- AmesHousing::make_ames()
```

More information on the dataset can be found here:
http://jse.amstat.org/v19n3/decock.pdf

Predicting Sale Price

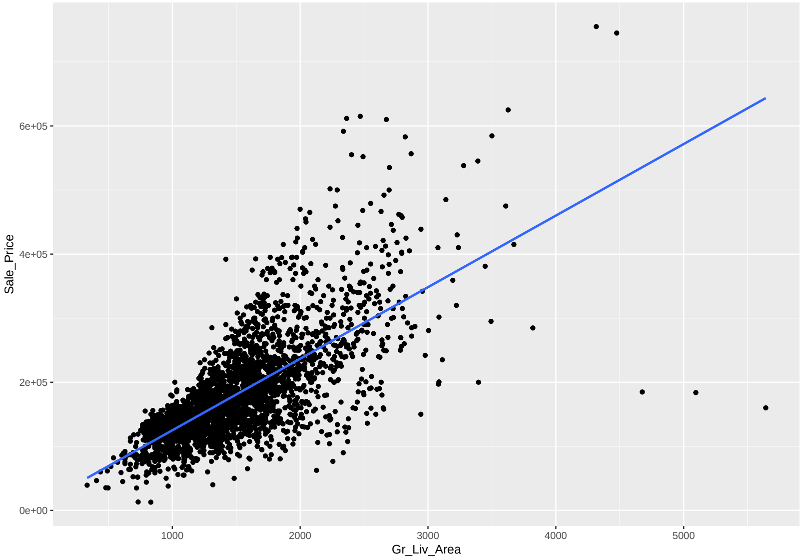
- Our goal in this section is to create a linear regression model that predicts the sale price of a home given its features:

$$\text{Sale\_Price} \sim f(\text{House Features})$$

- One feature that may be of importance is how big the house is, or Gr_Liv_Area:

$$\text{Sale\_Price} \sim f(\text{Gr\_Liv\_Area})$$

Plot of Sale_Price ~ Gr_Liv_Area



Linear model Sale_Price ~ Gr_Liv_Area

```
##
## Call:
## lm(formula = Sale_Price ~ Gr_Liv_Area, data = Ames)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -483467  -30219   -1966   22728  334323
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13289.634   3269.703   4.064 4.94e-05 ***
## Gr_Liv_Area   111.694      2.066  54.061  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
##
## Residual standard error: 56520 on 2928 degrees of freedom
## Multiple R-squared:  0.4995, Adjusted R-squared:  0.4994
```

Other house features

- We can do the same work for other house features that we think may be important, such as number of bedrooms:

$$\text{Sale\_Price} \sim f(\text{Bedroom\_AbvGr})$$

- or number of full bathrooms:

$$\text{Sale\_Price} \sim f(\text{Full\_Bath})$$

Multivariate Linear Regression

- Instead of building separate models, we can build one model that incorperates all variables:

$$\text{Sale\_Price} \sim f(\text{Gr\_Liv\_Area}, \text{Bedroom\_AbvGr}, \text{Full\_Bath})$$

- In general, a linear model with m predictors is given by:

$$Y \sim \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \cdots + \beta_m \cdot X_m$$

and is called **multivariate linear regression**.

Housing example

```
model3 <- lm(Sale_Price~Gr_Liv_Area+Bedroom_AbvGr, data=Ames)
summary(model3)
```

```
##
## Call:
## lm(formula = Sale_Price ~ Gr_Liv_Area + Bedroom_AbvGr, data
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -581397  -27840    -862   24526  329875
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    59496.236   3741.249   15.90   <2e-16 ***
## Gr_Liv_Area      136.361      2.247   60.69   <2e-16 ***
## Bedroom_AbvGr -29149.110   1372.135  -21.24   <2e-16 ***
## ---
```

Removing the intercept

- In some cases, it does not make sense to have the intercept in the model. For example, one would probably pay \$0 for a house that is 0 square feet, has 0 bedrooms and 0 bathrooms (ignoring plot of land):

```
model3 <- lm(Sale_Price~Gr_Liv_Area+Bedroom_AbvGr+0,
data=Ames)
```

Gradient descent for multivariate linear regression

- Gradient descent can be used for multivariate linear regression by adding the multiple X_k terms to the gradient.
- First we compute the cost:

$$C(\beta_0, \beta_1, \ldots, \beta_m) = \frac{1}{n}\sum_{i=1}^{n}(y_i - (\beta_0 + \beta_1 x_{i,1} + \cdots + \beta_m x_{i,m}))^2$$

Gradient descent for multivariate linear regression

- Next, we compute the gradient:

$$\nabla C = \begin{bmatrix} \dfrac{\partial C}{\partial \beta_0} \\ \dfrac{\partial C}{\partial \beta_1} \\ \vdots \\ \dfrac{\partial C}{\partial \beta_m} \end{bmatrix}.$$

where

$$\frac{\partial C}{\partial \beta_m} = -\frac{2}{n}\sum_{i=1}^{n} x_{i,k}(y_i - (\beta_0 + \beta_1 x_{i,1} + \cdots + \beta_m x_{i,m}))$$

Gradient descent for multivariate linear regression

- We still need the norm of the gradient, given by

$$||\nabla C|| = \left|\frac{\partial C}{\partial \beta_0}\right| + \left|\frac{\partial C}{\partial \beta_1}\right| + \cdots + \left|\frac{\partial C}{\partial \beta_m}\right|$$

  or

$$\sum_{i=0}^{m} \left|\frac{\partial C}{\partial \beta_i}\right|$$

Linear regression in vector notation

- Assume that Y and X_k are each vectors of length n:

$$X_k = \begin{bmatrix} x_{1,k} \\ x_{2,k} \\ \vdots \\ x_{n,k} \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},$$

for $k = 1, \ldots, m$.

Linear regression in vector notation

- Then the multivariate linear regression problem

$$Y \sim \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \cdots + \beta_m \cdot X_m$$

can be written in vector notation:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \sim \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \dots & x_{1,m} \\ 1 & x_{2,1} & x_{2,2} & \dots & x_{2,m} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n,1} & x_{n,2} & \dots & x_{n,m} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}$$

or

$$Y \sim X \cdot \mathbf{b}$$

where X is an $n \times (m+1)$ matrix containing the indicator variables and $\mathbf{b}$ is an $(m+1) \times 1$ column vector containing the unknown β_k values.

MSE in vector notation

- MSE, defined before as

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2.$$

can be written in matrix - vector notation:

$$MSE = \frac{1}{n}(Y - X\mathbf{b})^T(Y - X\mathbf{b})$$

where $(Y - X\mathbf{b})^T$ is the **transpose** of the vector $(Y - X\mathbf{b})$.

Gradient descent in vector notation

- The Cost function in gradient descent can be written as

$$C(\mathbf{b}) = \frac{1}{n}(Y - X\mathbf{b})^T(Y - X\mathbf{b})$$

- The gradient can be written as

$$\nabla C = -\frac{2}{n}X^T(Y - X\mathbf{b})$$

- Overall the algorithm becomes two lines after initializing:

while $(||\text{gradC}|| > \epsilon)\{$

$$\mathbf{b} = \mathbf{b} - \alpha * \text{gradC}$$

$$\text{gradC} = -\frac{2}{n}X^T(Y - X\mathbf{b})$$

$\}$