Machine Learning

author: Byteflow Dynamics date: 10/08/2017
incremental: true
class: small-code css: custom.css

Machine Learning

Machine learning refers to different methods of learning from data using a computer. There are two main different types of machine learning: Supervised and Unsupervised.

Supervised Learning Supervised learning is basically pattern recognition. You train your algorithm with a set of *labeled* data, then the trained algorithm will be used to predict labels for new data points.

Majority of machine learning problems are supervised learning, and they can be:

- Classification
- Regression

Unsupervised Learning

Unsupervised learning is used when data is not labeled. It's used for data exploration to find patterns or structure in the data. The most common types are:

- Principal Component Analysis
- Clustering

Case Study: Advertising

- Data on money spent on several advertising outlets and volume sales
  (Data can be found here: http://www-bcf.usc.edu/~gareth/ISL/data.html)

```
  X    TV radio newspaper sales
1 1 230.1  37.8      69.2  22.1
2 2  44.5  39.3      45.1  10.4
3 3  17.2  45.9      69.3   9.3
4 4 151.5  41.3      58.5  18.5
5 5 180.8  10.8      58.4  12.9
6 6   8.7  48.9      75.0   7.2
```

Possible Questions to ask about the data as a Consultant

- Is there a relationship between advertising budget and sales?
- How strong is the relationship between advertising budget and sales?
- Which media contribute to sales?
- How accurately can we estimate the effect of each medium on sales?

- How accurately can we predict future sales?

- Is the relationship linear?

Simple Linear Models

- Predict sales based on TV, Radio and Newspaper ads
- Assume a linear model with one variable.

$$Y = \beta_0 + \beta_1 X + \epsilon$$

* Where the unknown coefficients (parameters) are β_0 and β_1, y-intercept and slope, ϵ is the error term.
* Use the **training data** to calculate the coefficients:

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x$$

* $\widehat{y}$ -prediction of Y

Advertising, TV

Assume a simple linear relationship between TV spending and Sales:

$$sales = \beta_0 + \beta_1 * TV$$

Lets look at the data

lm()

class: small-code

- Use the built in lm() function to calculate the coefficients

```
lm.fit <- lm(sales ~ TV, data = Advertising)
summary(lm.fit)
```

```

Call:
lm(formula = sales ~ TV, data = Advertising)

Residuals:
    Min      1Q  Median      3Q     Max
-8.3860 -1.9545 -0.1913  2.0671  7.2124

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 7.032594   0.457843   15.36   <2e-16 ***
TV          0.047537   0.002691   17.67   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.259 on 198 degrees of freedom
Multiple R-squared:  0.6119,    Adjusted R-squared:  0.6099
F-statistic: 312.1 on 1 and 198 DF,  p-value: < 2.2e-16
```

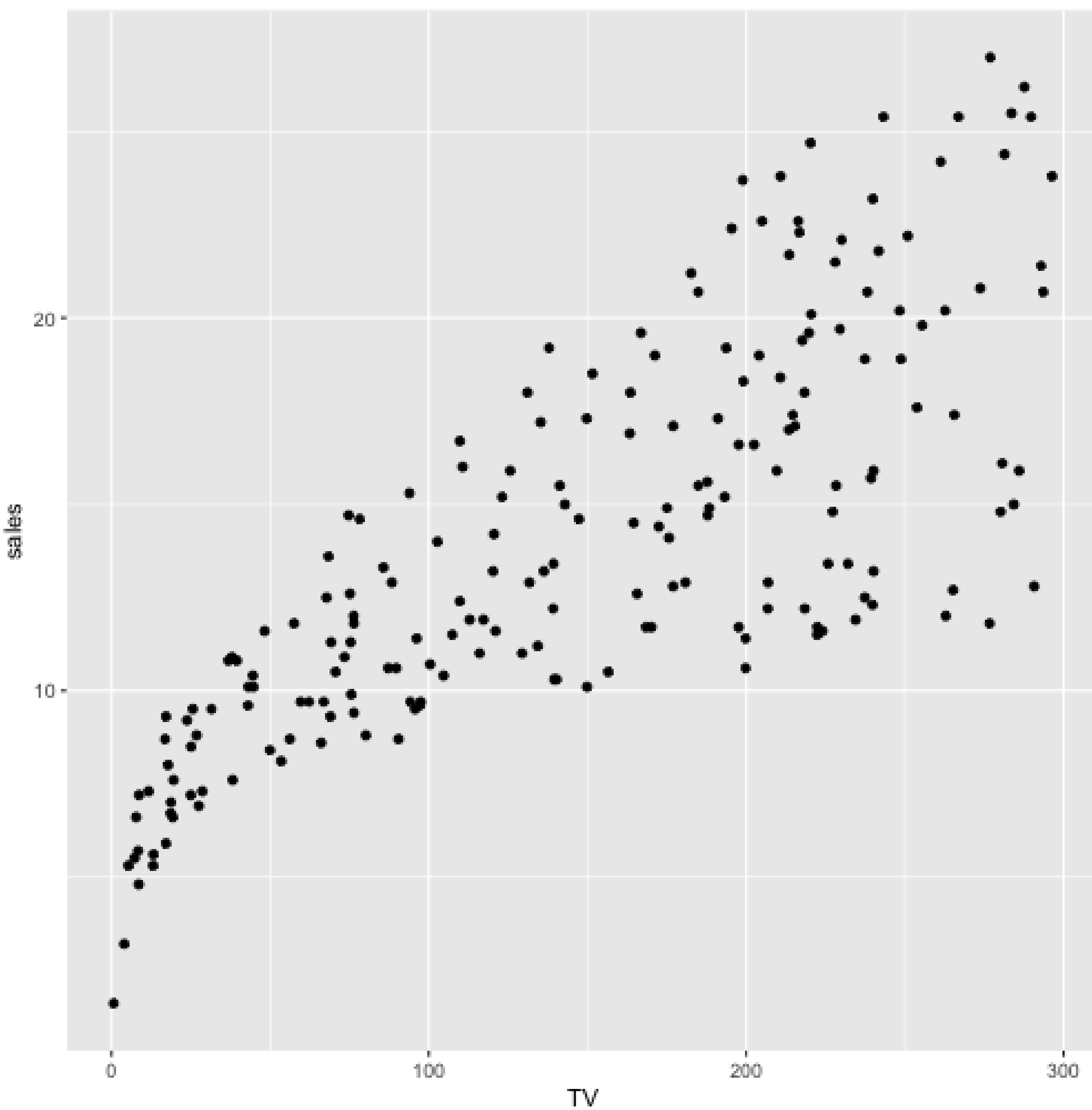

Figure 1: plot of chunk unnamed-chunk-1

Calculating the coefficient

- lm( ) calculates the slope and y-intercept
- How?
  let y_i be the prediction of Y based on x_i
  $\widehat{y_i} = \widehat{\beta}_0 + \widehat{\beta}_1 x_i$
  let $e_i = y_i - \widehat{y_i}$ be the i-th residual (the difference between i-th observed response and i-th predicted value by our model)
- Minimize the Residual Sum of Squares (RSS)

$$RSS = e_1^2 + e_2^2 + ... + e_n^2$$
$$RSS = (y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_1)^2 + ... + (y_n - \widehat{\beta}_0 - \widehat{\beta}_1 x_n)^2$$

Assessing the accuracy of the model

Standard error of the slope:

$$SE(\widehat{\beta}_1)^2 = \frac{\sigma^2}{\Sigma_{i=1}^n (x_i - \bar{x})}$$

where $\sigma^2 = Variance(\epsilon)$

- The wider the x-space the lower the Standard Error
- Training data should be random and cover the entire spectrum

Confidence Intervals

- SE can be used to compute confidence intervals of our coefficients (if the error distribution is Gaussian).
- 95% confidence interval of the slope $\widehat{\beta}_1 \pm 2 * SE(\widehat{\beta}_1)$
- 95% chance that the interval contains the true value of β_1
- $\widehat{\beta}_1$ is an approximation, calculated from the one sample data.
- All this is done automatically in R.

t-statistic

- Is there a significant relationship between X and Y? If not the slope is zero.
- Is the slope far enough from zero?
- First calculate the standard error $SE(\hat{\beta}_1)$.
- If $SE(\widehat{\beta}_1)$ is small, we can reject the null hypothesis.
- If $SE(\widehat{\beta}_1)$ is large then $\widehat{\beta}_1$ must be large.

t-statistic

- Calculate the t-statistic: a measure of the number of standard deviations that $\hat{\beta}_1$ is away from zero.

$$t = \frac{\widehat{\beta}_1 - 0}{SE(\widehat{\beta}_1)}$$

.

- p-value: the probability of observing any value bigger then $|t|$, if $\beta_1 = 0$.
- Typical p-values cutoff for rejecting null-hypothesis are 5% or 1%.

lm.fit

class: small-code * Lets take a look at the coefficients from the advertising data.

```r
summary(lm.fit)
```

```

Call:
lm(formula = sales ~ TV, data = Advertising)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.3860 -1.9545 -0.1913  2.0671  7.2124 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 7.032594   0.457843   15.36   <2e-16 ***
TV          0.047537   0.002691   17.67   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.259 on 198 degrees of freedom
Multiple R-squared:  0.6119,    Adjusted R-squared:  0.6099 
F-statistic: 312.1 on 1 and 198 DF,  p-value: < 2.2e-16
```

R^2 Statistic

- Explains proportion of the varience. Takes value between 0 an 1.
- To calculare R^2 use:

$$R^2 = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}$$

where $TSS = \Sigma(y_i - \bar{y})^2$ is the *total sum of squares*

Newspaper realtionship

- Excercise:

1. Calculate the slope of sales vs. newspaper
2. Does newspaper spending have an effect on sales? What effect does spending $1000 have on sales?
3. Plot sales vs. newspaper

Multiple Linear Regression

class: small-code * Linear fit of all the parameters.

```r
lm.fit <- lm(sales ~ ., data = Advertising)
summary(lm.fit)
```

```

Call:
lm(formula = sales ~ ., data = Advertising)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.8105 -0.9008  0.2641  1.1783  2.8336 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  3.0052094  0.3942082   7.623 1.06e-12 ***
X           -0.0005798  0.0020992  -0.276    0.783    
TV           0.0457759  0.0013988  32.725  < 2e-16 ***
radio        0.1883832  0.0086480  21.784  < 2e-16 ***
newspaper   -0.0012433  0.0059319  -0.210    0.834    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.689 on 195 degrees of freedom
Multiple R-squared:  0.8973,	Adjusted R-squared:  0.8951 
F-statistic: 425.7 on 4 and 195 DF,  p-value: < 2.2e-16
```

Multiple Linear Regression

- Newspaper advertising seem to have no effect.
- ..while the TV and Radio spending is at current rate.
- It probably does have an effect if TV and Radio spending is zero.
- To study this effect the system must be purterbed.