Multiple Regression

Concept Module 12

Template by: Laurent Lessard

Review: Regression with one feature

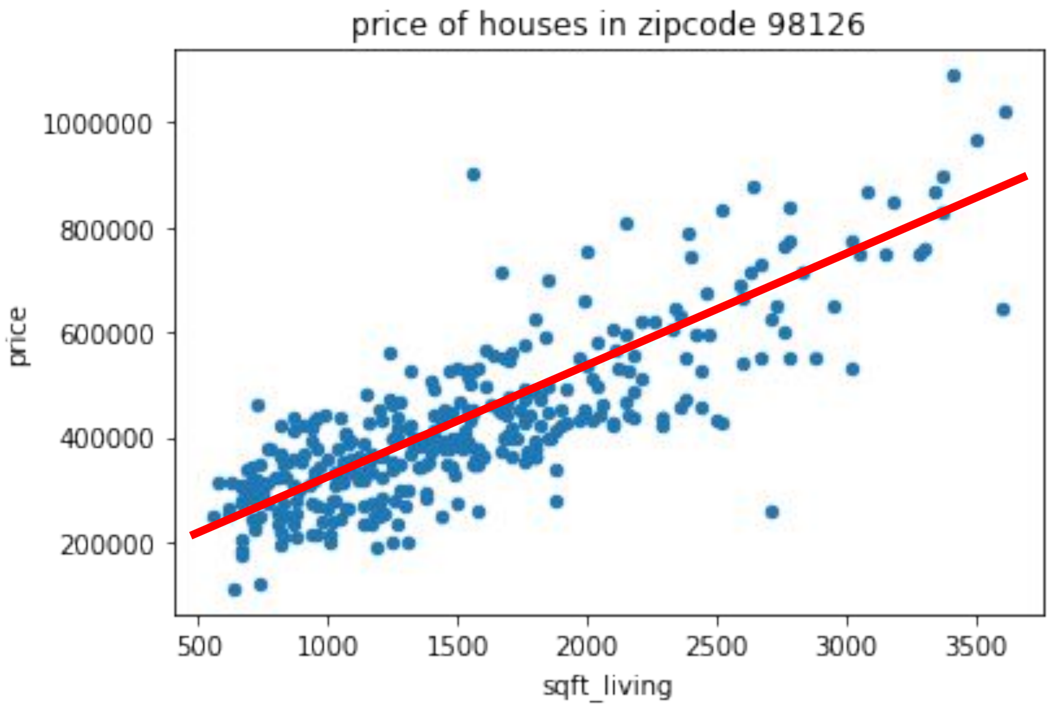
Main idea: find a line that best fits the data

Equation of line:

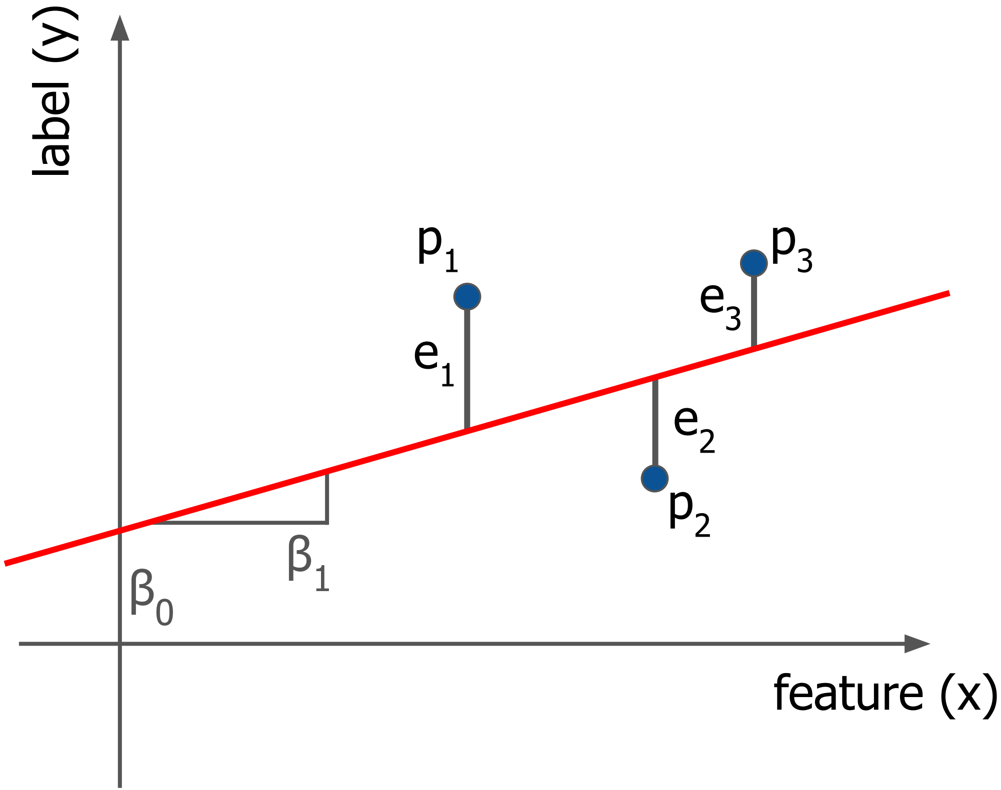
$$(\text{price}) = \beta_0 + \beta_1(\text{sqft\_living})$$

intercept

slope



Review: Geometry of linear regression



Goal: choose β_0, β_1 to minimize the sum of squared residuals:

$$RSS = (e_1)^2 + \ldots + (e_n)^2$$

Multiple regression

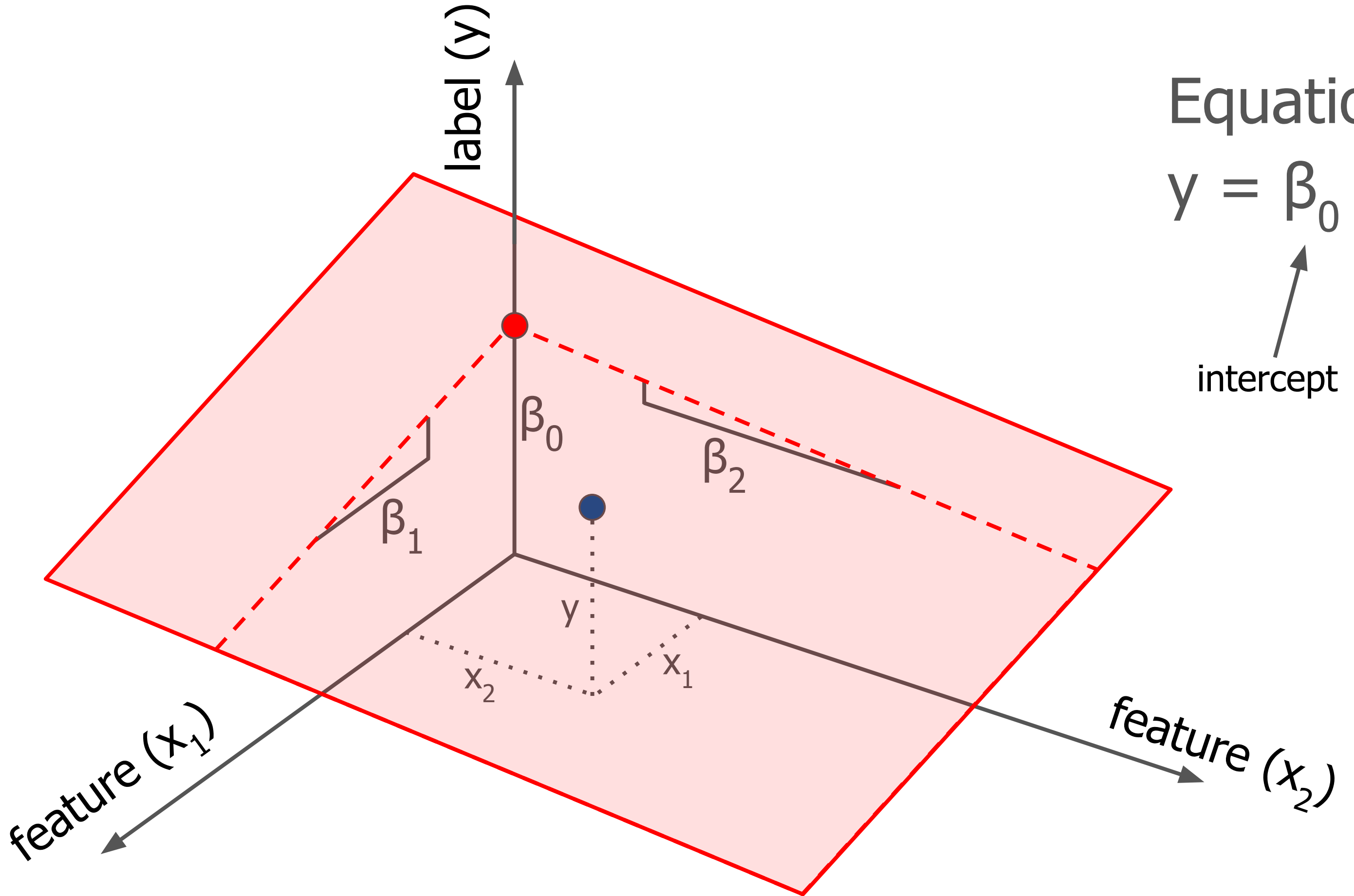
Main idea: fit the data using a linear combination of features

Linear combination using intercept + two features:

$$(\text{price}) = \underbrace{\beta_0}_{\text{intercept}} + \beta_1(\text{sqft\_living}) + \beta_2(\text{sqft\_lot})$$

intercept slopes

Geometry of multiple regression



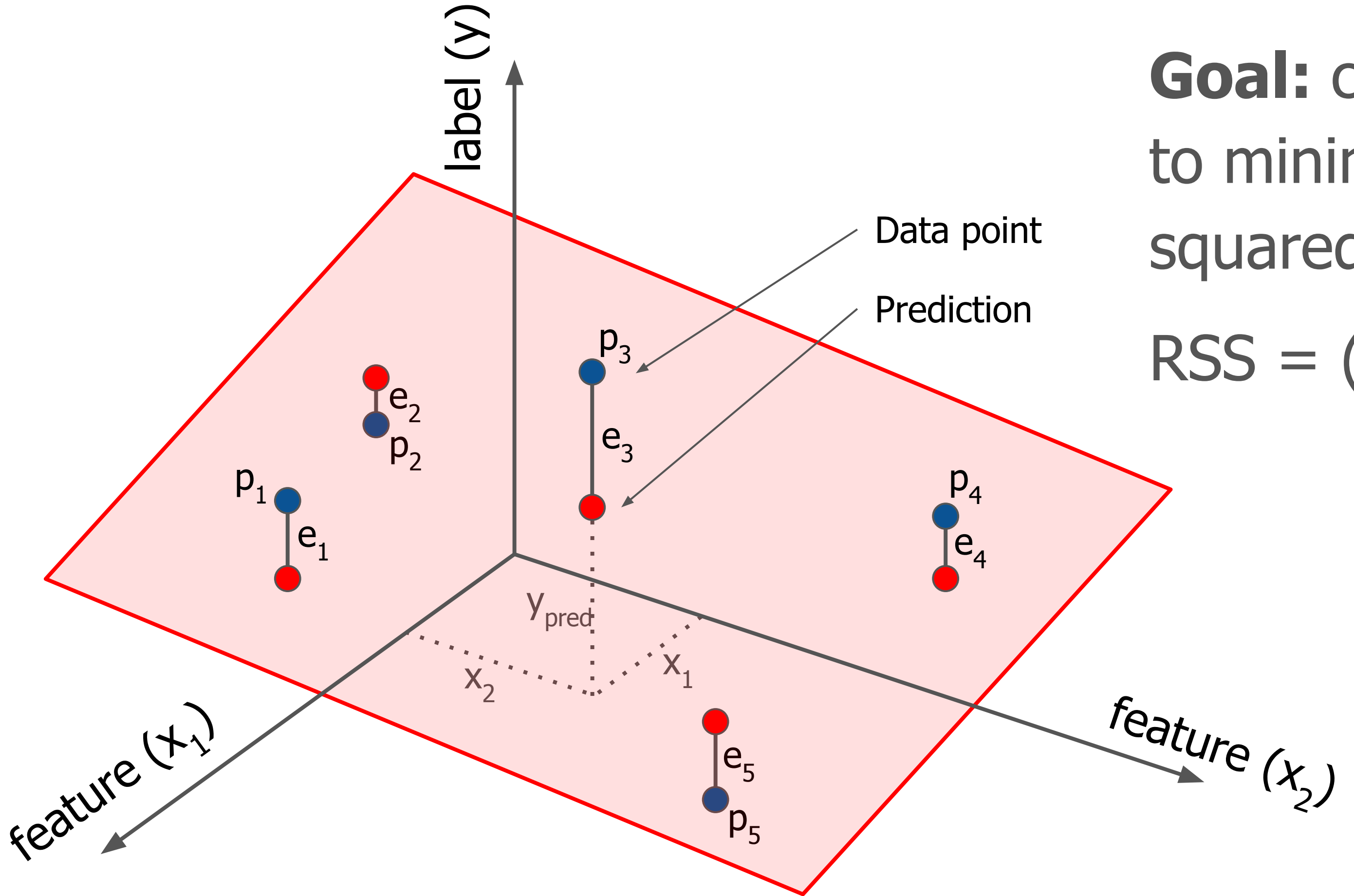
Equation of plane:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

intercept

slopes

Geometry of multiple regression



Goal: choose β_0, β_1, β_2 to minimize the sum of squared residuals:

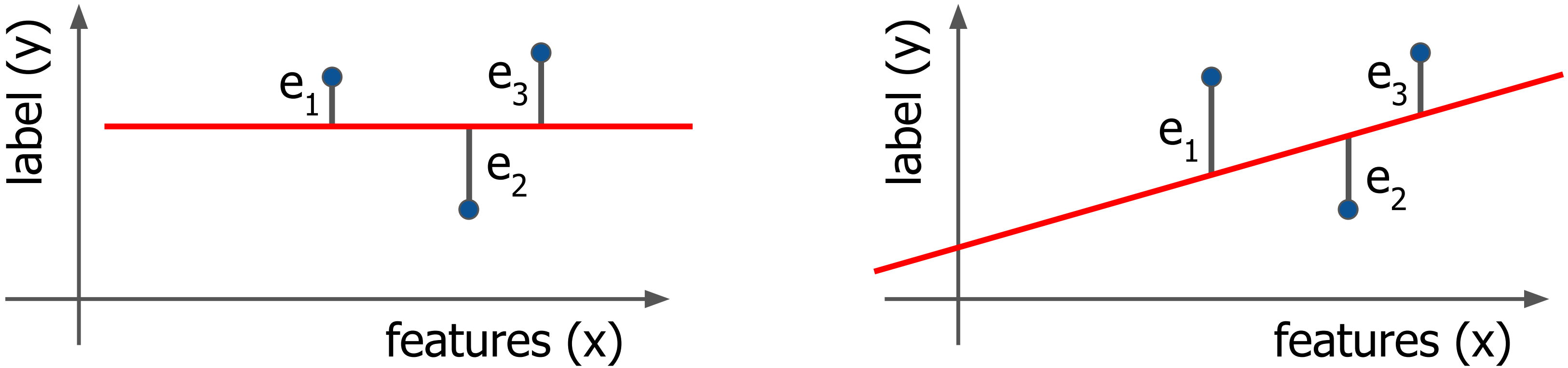
$$RSS = (e_1)^2 + \dots + (e_n)^2$$

Total vs Residual sum of squares

If we force $\beta_1 = \beta_2 = \ldots = \beta_m = 0$ (no slopes), then the best we can do is to set β_0 = mean(y). Then RSS = variance of y. We call this the "total sum of squares" (TSS).

TSS = total variance of the labels

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



Array-vector multiplication

Linear model:

$$y \approx \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m$$

This holds for each example:

$$y^{(0)} \approx \beta_0 + \beta_1 x_1^{(0)} + \beta_2 x_2^{(0)} + \dots + \beta_m x_m^{(0)}$$

$$y^{(1)} \approx \beta_0 + \beta_1 x_1^{(1)} + \beta_2 x_2^{(1)} + \dots + \beta_m x_m^{(1)}$$

$$\vdots$$

$$y^{(n)} \approx \beta_0 + \beta_1 x_1^{(n)} + \beta_2 x_2^{(n)} + \dots + \beta_m x_m^{(n)}$$

Array-vector multiplication

$$\begin{pmatrix} y^{(0)} \\ y^{(1)} \\ \vdots \\ y^{(n)} \end{pmatrix} \approx \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \beta_0 + \begin{pmatrix} x_1^{(0)} \\ x_1^{(1)} \\ \vdots \\ x_1^{(n)} \end{pmatrix} \beta_1 + \begin{pmatrix} x_2^{(0)} \\ x_2^{(1)} \\ \vdots \\ x_2^{(n)} \end{pmatrix} \beta_2 + \ldots + \begin{pmatrix} x_m^{(0)} \\ x_m^{(1)} \\ \vdots \\ x_m^{(n)} \end{pmatrix} \beta_m$$

Array-vector multiplication

$$\begin{pmatrix} y^{(0)} \\ y^{(1)} \\ \vdots \\ y^{(n)} \end{pmatrix} \approx \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \beta_0 + \begin{pmatrix} x_1^{(0)} & x_2^{(0)} & \dots & x_m^{(0)} \\ x_1^{(1)} & x_2^{(1)} & \dots & x_m^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(n)} & x_2^{(n)} & \dots & x_m^{(n)} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{pmatrix}$$

$$\boxed{y \approx \beta_0 + X \beta}$$

Multiple linear regression in Python

```
from sklearn.linear_model import LinearRegression

X = df[['sqft_living','sqft_lot']]  # features
y = df['price']                     # labels

regr = LinearRegression()
regr.fit(X,y)
```

Slope and intercept

```
# Obtain intercept
regr.intercept_

# Obtain slope
regr.coef_
```

Predict new data

```
# Predict labels for
# new unlabeled data
ytest = regr.predict(Xtest)
```

Get R^2 score

```
# Obtain R-squared
regr.score(X,y)
```

Housing data result

Using 'sqft_living' only

```
# Obtain intercept
regr.intercept_
```

123936.34087573813

```
# Obtain slope
regr.coef_
```

array([194.47792472])

```
# Obtain R-squared
regr.score(X,y)
```

0.6574621635804526

Using 'sqft_living' and 'sqft_lot'

```
# Obtain intercept
regr.intercept_
```

91339.341146513645

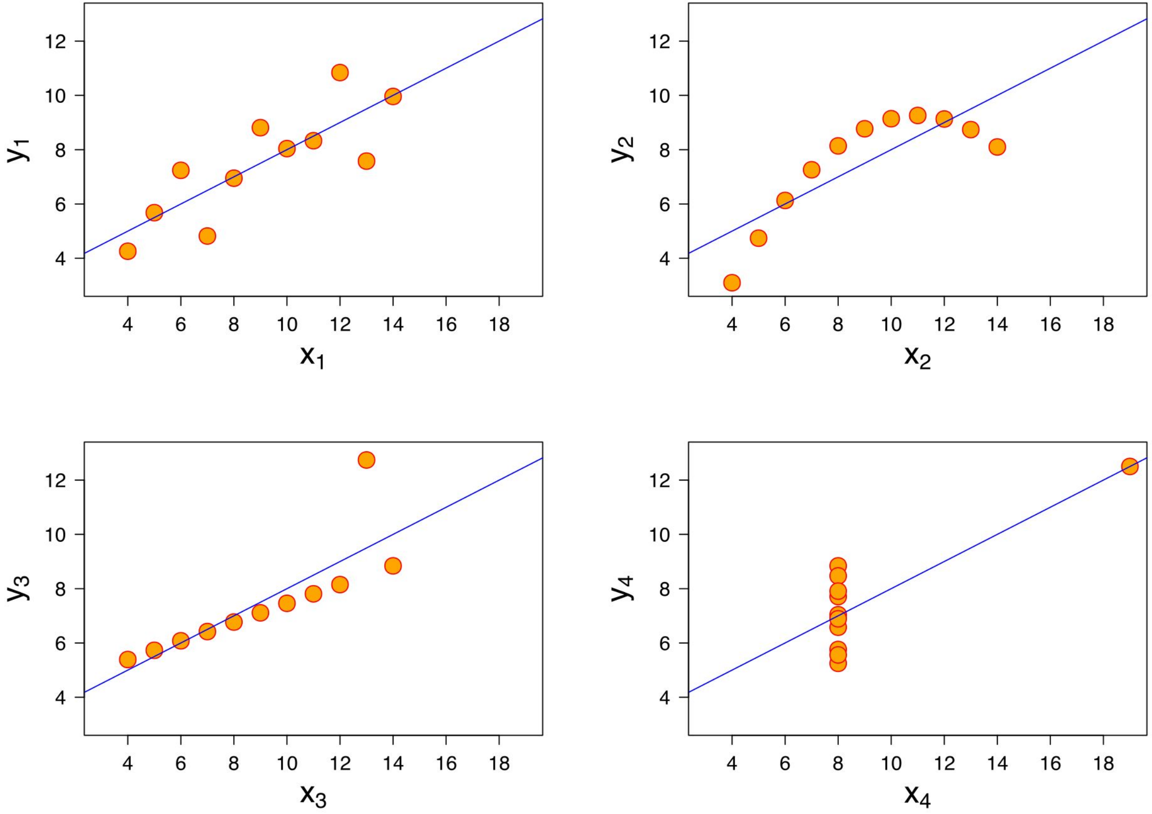
```
# Obtain slope
regr.coef_
```

array([191.91219356, 7.16840722])

```
# Obtain R-squared
regr.score(X,y)
```

0.6654997011537487

Warning: R^2 isn't everything!

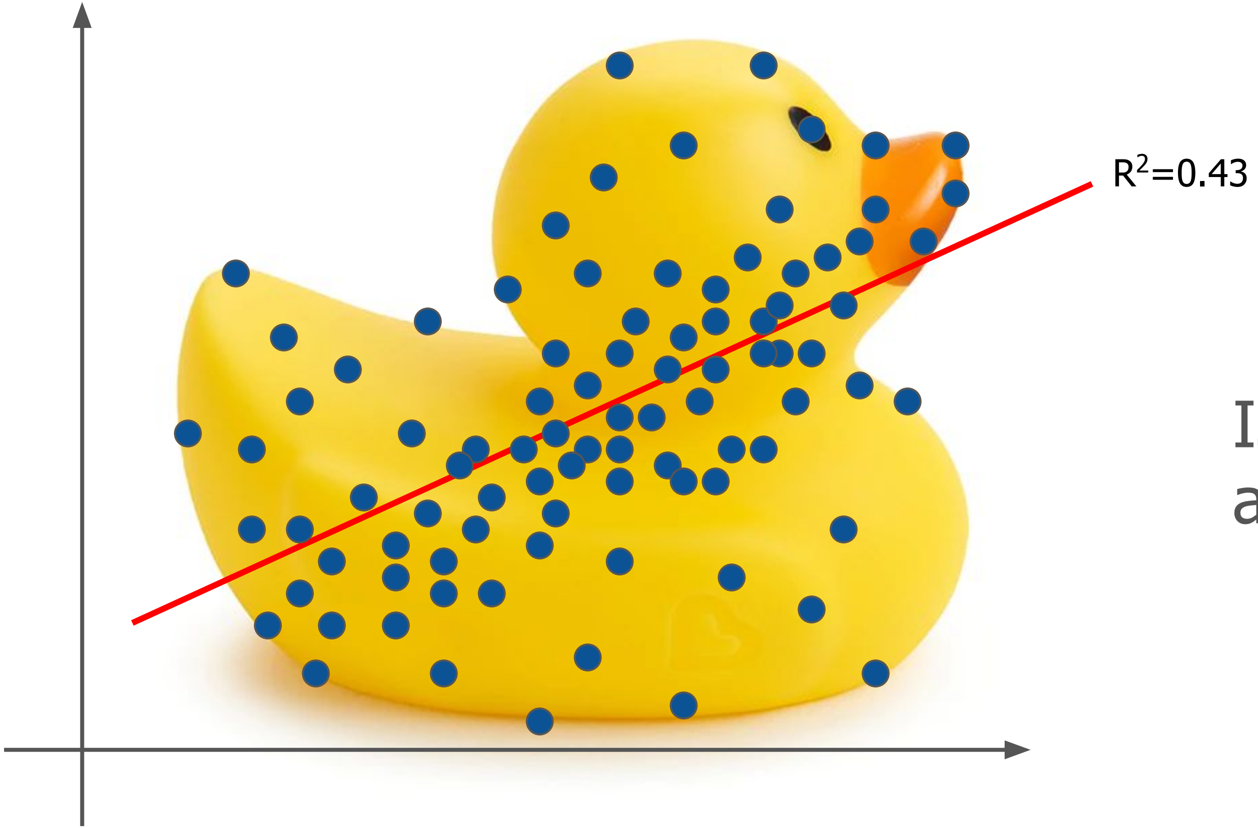


These four pairs of points have:

- same mean(x) = 9
- same std(x) = 3.32
- same mean(y) = 7.5
- same std(y) = 2.03
- same line of best fit: y=3+0.5x
- same R^2=0.67

source: https://en.wikipedia.org/wiki/Anscombe%27s_quartet

Warning: Is your data a duck?



Is regression always a good idea?

Summary

- Multiple regression fits the label to a linear combination of the features (just like array-vector multiplication!)
- Surface is characterized by slope(s) and intercept.
- Linear regression minimizes sum of squared residuals (RSS)
- Be careful: Is your data a duck?