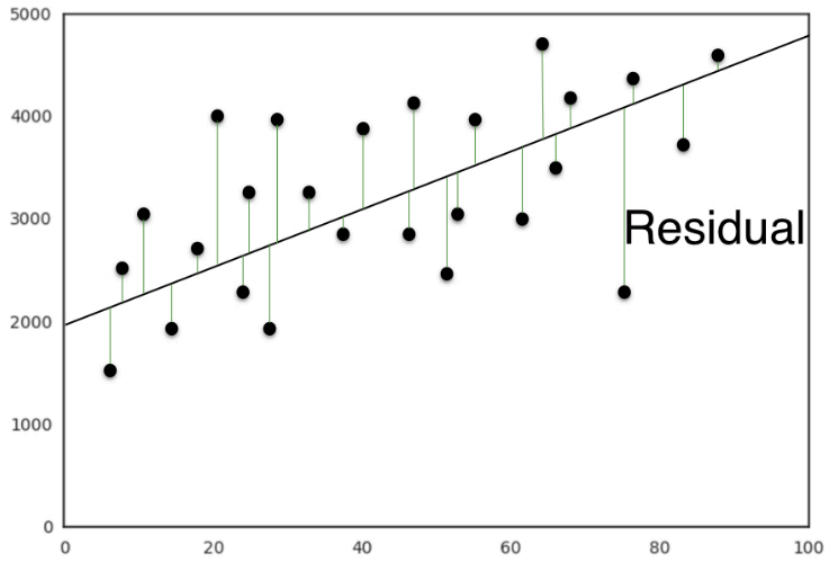
Linear Regression

Regression mechanics

- y = ax + b
  - y = target
  - x = single feature
  - a, b = parameters of model
- How do we choose a and b?
- Define an error functions for any given line
  - Choose the line that minimizes the error function

The loss function



Residual Squared Error, because we do not want the positive and negative

residuals to cancel out each other out,

Linear regression in higher dimensions

$$y = a_1x_1 + a_2x_2 + b$$

- To fit a linear regression model here:
  - Need to specify 3 variables
- In higher dimensions:
  - Must specify coefficient for each feature and the variable b

$$y = a_1x_1 + a_2x_2 + a_3x_3 + ... + a_nx_n + b$$

- Scikit-learn API works exactly the same way:
  - Pass two arrays: Features, and target

Linear regression on all features

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression

X_train, X_test, y_train, y_test = train_test_split(X, y,
    test_size = 0.3, random_state=42)
reg_all = LinearRegression()
reg_all.fit(X_train, y_train)
y_pred = reg_all.predict(X_test)
reg_all.score(X_test, y_test)
```

```
0.71122600574849526
```