

ROOT MEAN SQUARED ERROR

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (p_i - y_i)^2}{n}}$$

Where p is the predicted value,
 y is the actual value,
 i is the index,
for a particular data point.
 n data points.

GRADIENT DESCENT

A WIDELY USED OPTIMIZATION ALGORITHM

- Optimize a function by following the gradient of the cost function.
- Not widely used for Linear Regression unless the data set is very large (cannot fit in memory to perform OLS regression).
- Used behind the scenes in a wide variety of model training operations in Python's scikit-learn.
- It's a so-called "on-line" algorithm as it processes and learns from the data one sample at a time. (Memory Efficient)
- Stochastic Gradient Descent is a variant where the data is shuffled occasionally to avoid stagnation.

STOCHASTIC GRADIENT DESCENT

PSEUDO-CODE (WE'LL IMPLEMENT THIS IN PYTHON)

```
assume input vector x (age) and vector y (height)

#we want to predict for any age the weight of the sample accurately

coefficients B0=0.0 and B1=0.0, model:  $y=B0+B1x$ 

while(rmse > error_threshold && epochs < max_epochs):

    shuffle the data

    for each data point i:

        error = model_prediction(x[i]) - y[i]

        B0 = B0 - learning_rate*error

        B1 = B1 - learning_rate*error*x[i]

    epochs = epochs + 1

    compute rmse
```