**Statistics for Machine Learning:**

**Population vs. Sample**

Example: Average height of all the people in a state

Population= all the people in a state

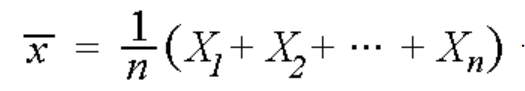
Population Mean = sum of (let’s say) 1 million heights/1 million

The formula to find the population mean is:  
μ = (Σ \* X)/ N  
where:  
Σ means “the sum of.”  
X = all the individual items in the group.  
N = the number of items in the group.

Usually for estimation of such large datasets, we use a sample:

Sample= subset of the population (say, around 10,000)

The ***sample mean*** http://www.stat.yale.edu/Courses/1997-98/101/xbar.gif from a group of observations is an estimate of the population mean http://www.stat.yale.edu/Courses/1997-98/101/mu.gif. (As the sample size increases, the sample mean will almost be equal to the population mean)



Now, the sample mean can be easily computed = sum of 10,000 height/10,000