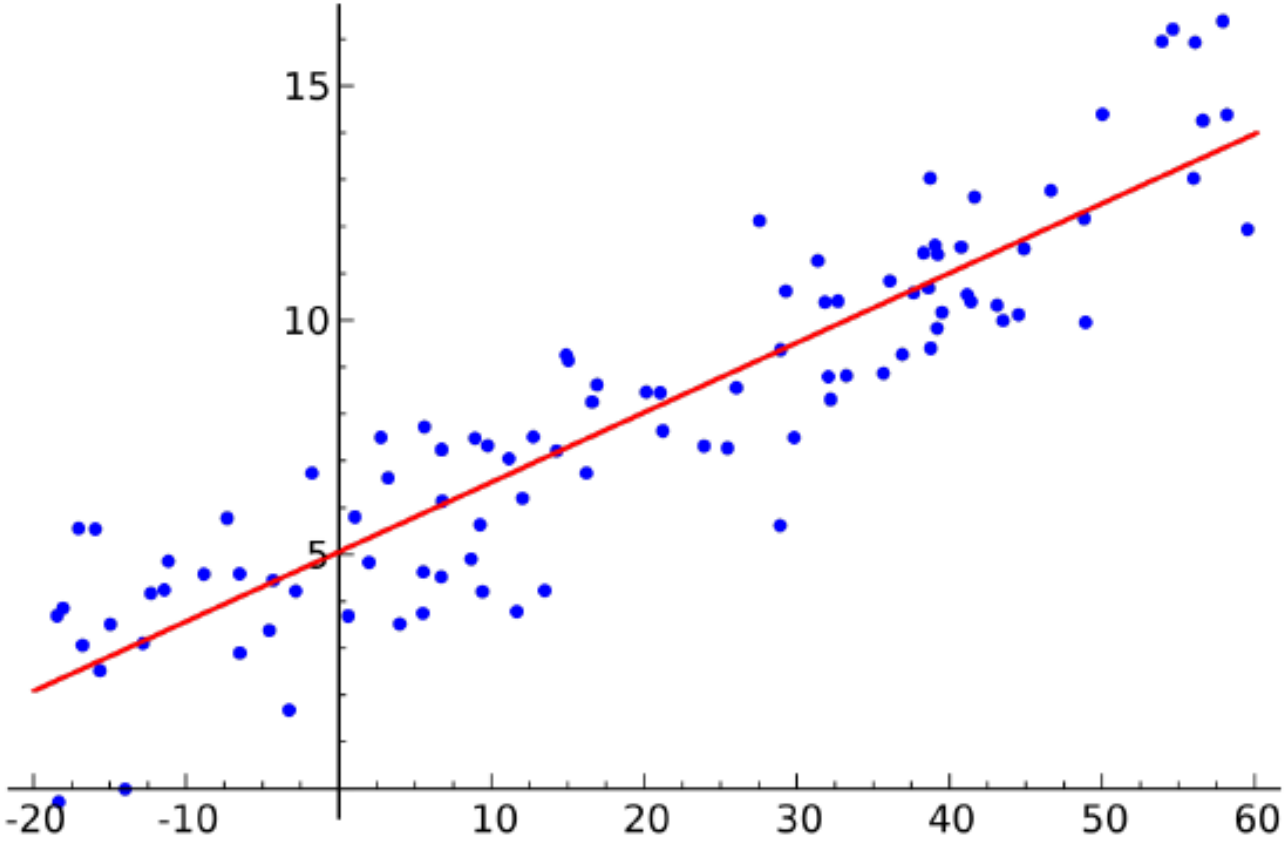
EE 595 Lab6

Linear Regression

- Linear Regression tries to fit a model which is linear/affine



Linear Regression

- Original Data: $\underline{x} = [x_1, x_2, \ldots, x_d]$, $\underline{y}$
- Augmented Data: $\underline{x}' = [1, x_1, x_2, \ldots, x_d]$
- Hypothesis:

$$h_w(x) = \underline{w}^T \underline{x}' = w_0 + w_1 x_1 + w_1 x_1 + \cdots + w_d x_d$$

- Cost Function:

$$J(w_0, w_1, \ldots, w_d) = \frac{1}{2N} \sum_{i=1}^{N} (h_w(x^{(i)}) - y^{(i)})^2$$

- Goal: Find w that minimize $J(w)$

K-Means clustering

- Partition N data points into K clusters
- Steps:
  - **Step 0:** obtain an initial set of k means (k centers)
  - **Step 1:** assign data points to the cluster whose mean has the least square Euclidean distance (“nearest”)

$$S_i^{(t)} = \left\{ x_p : \left\| x_p - m_i^{(t)} \right\|^2 \leq \left\| x_p - m_j^{(t)} \right\|^2 \ \forall j, 1 \leq j \leq k \right\},$$

  - **Step 2:** calculate the new means (center) of the data points in the new clusters

$$m_i^{(t+1)} = \frac{1}{\left| S_i^{(t)} \right|} \sum_{x_j \in S_i^{(t)}} x_j$$

  - Step 3: Determine whether to stop or return to step 1

Example