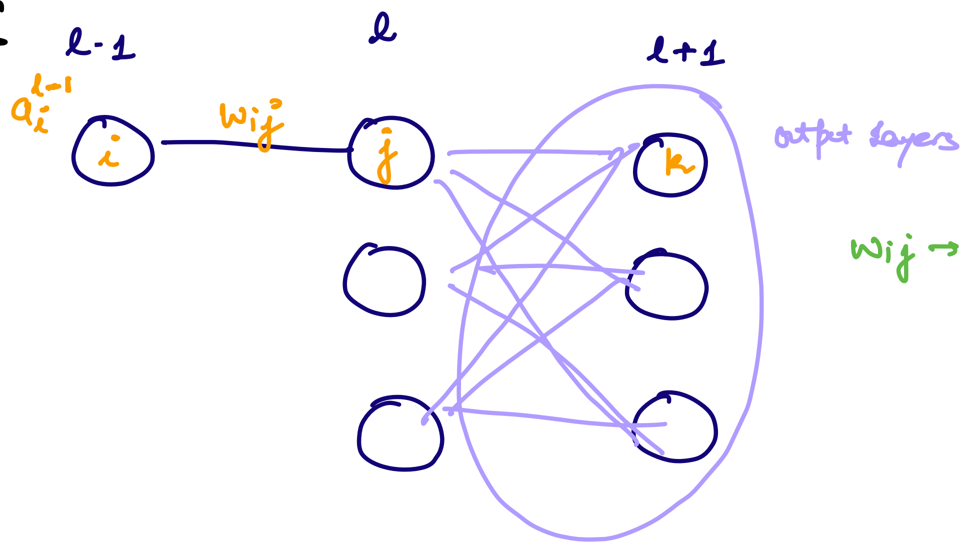
Hidden Layer



$w_{ij} \to z_j \to a_j \to z_k \to a_k \to L = (y - a_k)$

$$\frac{\partial L}{\partial w_{ij}^{l}} = \frac{\partial L}{\partial z_k^{l+1}} \cdot \frac{\partial z_k^{l+1}}{\partial a_j^{l}} \cdot \frac{\partial a_j^{l}}{\partial z_j^{l}} \cdot \frac{\partial z_j^{l}}{\partial w_{ij}^{l}}$$

$$x = x - \eta \frac{\partial L}{\partial x}$$

$$\frac{\partial z_k^{l+1}}{\partial a_j^{l}}$$

$$z_k^{l+1} = \sum_j w_{jk} a_j^{l} + b^{l+1}$$

$$\frac{\partial z_k^{l+1}}{\partial a_j^{l}} = w_{jk}$$

$$\frac{\partial a_j^{l}}{\partial z_j^{l}}$$

$$a_j = \sigma(z_j)$$

$$\frac{\partial a_j}{\partial z_j} = \sigma'(z_j)$$

$$\frac{\partial z_j^{l}}{\partial w_{ij}^{l}}$$

$$z_j = w_{ij} a_i^{l-1} + b_j^{l}$$

$$\frac{\partial z_j}{\partial w_{ij}} = a_i^{l-1}$$

$$\frac{\partial L}{\partial w_{ij}^l} = \overbrace{\underbrace{\frac{\partial L}{\partial z_k^{l+1}}}_{\delta_k^{l+1}} \cdot \frac{\partial z_k^{l+1}}{\partial a_j^l} \cdot \frac{\partial a_j^l}{\partial z_j^l}}^{\delta_j^l} \cdot \frac{\partial z_j^l}{\partial w_{ij}^l}$$

$$= \underbrace{\delta_k^{l+1} \quad w_{jk}^{l+1} \quad \sigma'(z_j^l)}_{\delta_j^l} \quad a_i^{l-1}$$

$$\frac{\partial L}{\partial w_{ij}^l} = \delta_j^l \cdot a_i^{l-1}$$

Bias:

$$\frac{\partial L}{\partial b_j^l} = \overbrace{\underbrace{\frac{\partial L}{\partial z_k^{l+1}}}_{\delta_k^{l+1}} \cdot \frac{\partial z_k^{l+1}}{\partial a_j^l} \cdot \frac{\partial a_j^l}{\partial z_j^l}}^{\delta_j^l} \cdot \frac{\partial z_j^l}{\partial b_j^l} \longrightarrow 1$$

$$\frac{\partial L}{\partial b_j^l} = \delta_j^l \cdot 1$$

$$z_j = \underbrace{w_{ij} \; a_i^{l-1}} + \underbrace{b_j^l}$$

$$\frac{\partial z_j}{\partial b} = 0 + 1$$

Final Results:

$$\frac{\partial L}{\partial w_{ij}^l} = \delta_j^l \cdot a_i^{l-1}$$

$$\frac{\partial L}{\partial b_j^l} = \delta_j^l \cdot 1$$

$$\delta_k^{l+1} \quad w_{jk}^{l+1} \quad \sigma'(z_j^l)$$

δ_j^{l}

$$w_{ij} = w_{ij} - \eta \frac{\partial L}{\partial w_{ij}}$$

$$b_j = b_j - \eta \frac{\partial L}{\partial b_j}$$

Gradient

Loss fx^n: Binary Cross Entropy (Binary Classification)

$$L = -\left(y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)\right)$$

$\hat{y}_i$ → activation value in o/p layer $= \hat{y}$

Output:

$$\delta_j = \frac{\partial L}{\partial z_j} = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial z}$$

$a \rightarrow \hat{y}$

$$= \left(-\frac{y_i}{\hat{y}_i} + \frac{(1-y_i)}{(1-\hat{y}_i)}\right)$$

$$a = \sigma(z)$$

$$\frac{\partial a}{\partial z} = \sigma(z)(1-\sigma(z))$$

$$= a(1-a)$$

$$= \hat{y}(1-\hat{y})$$

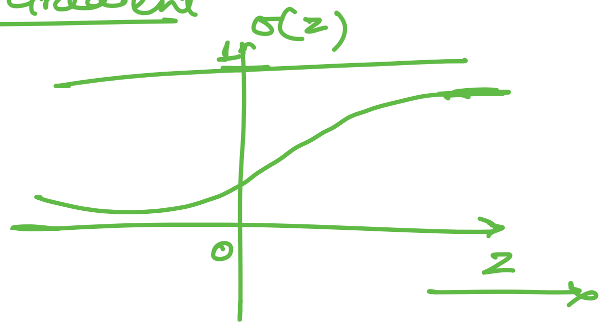
$$\delta_j = \left(\frac{-y_i}{\hat{y}_i} + \frac{1-y_i}{1-\hat{y}_i}\right)\left(\hat{y}_i(1-\hat{y}_i)\right)$$

$$= \left(\frac{-y_i + \cancel{y_i\hat{y}_i} + \hat{y}_i - \cancel{y_i\hat{y}_i}}{\cancel{\hat{y}_i(1-\hat{y}_i)}}\right)\cancel{\left(\hat{y}_i(1-\hat{y}_i)\right)}$$

$= (\hat{y}_i - y_i)$

Vanishing Gradient

sigmoid



$$\sigma'(z) = \sigma(z)(1-\sigma(z))$$

z is large

$$\sigma'(z) = \sigma(z)\underbrace{(1-\underbrace{\sigma(z)}_{1})}_{0}$$

$$\sigma'(z) = 0$$

z is small

$$\sigma'(z) = \underbrace{\sigma(z)}_{0}(1-\sigma(z))$$

$$= 0$$

$$w = w - \eta \frac{\partial L}{\partial w} \rightarrow 0$$

no change in wts.

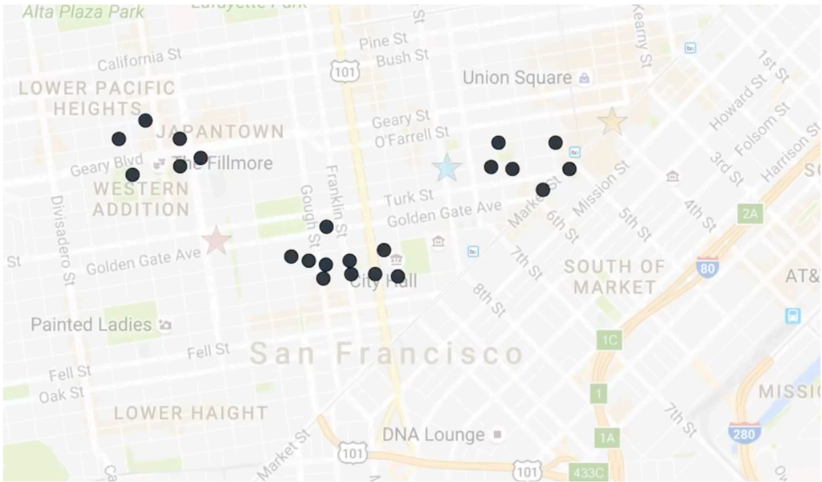
Supervised Learning Algo
- Regression (Linear Regression)
- Classification
  - Logistic
  - NB
  - NN
  - DT
  - KNN

Unsupervised Algorithms:

x features, y ✗

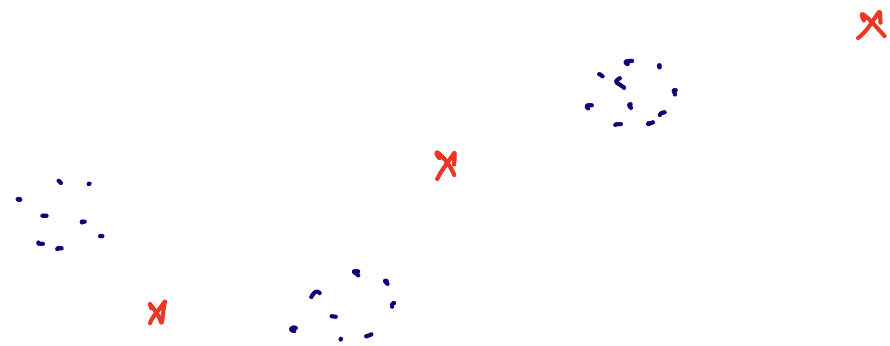
Clustering

group similar kind of data together.

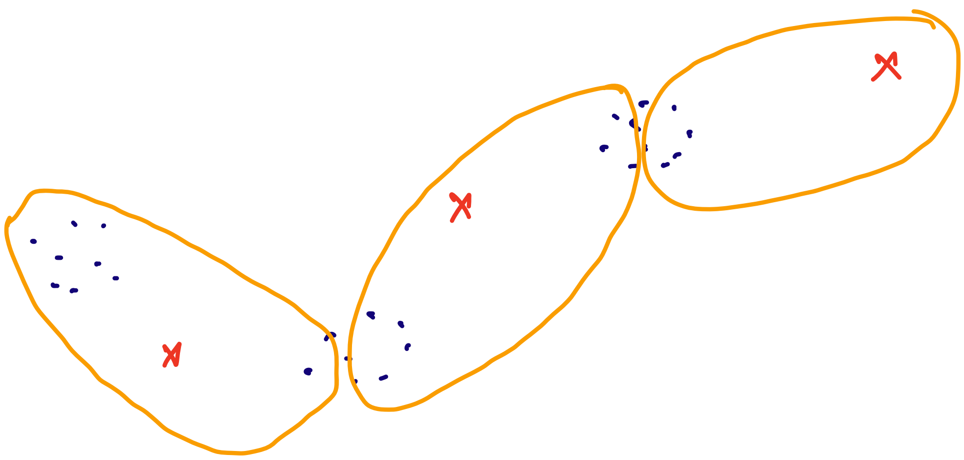


3 outlets ?

1. Randomly initialize 3 center points

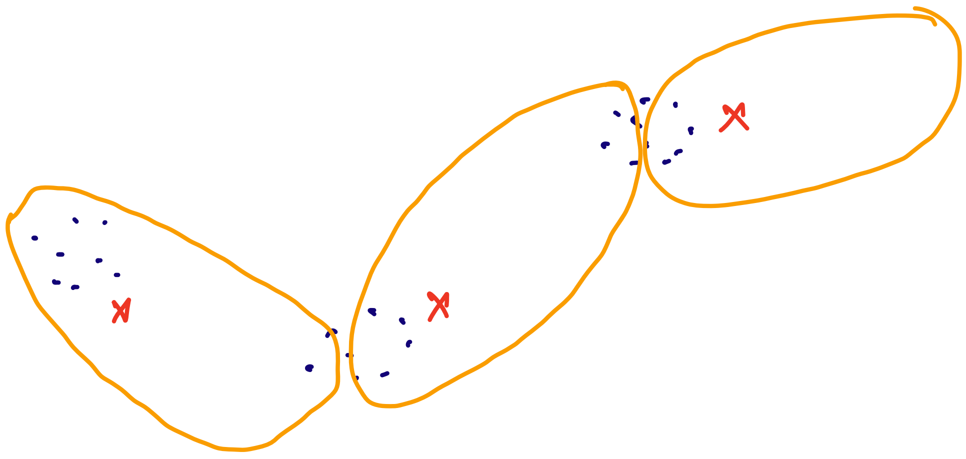


2. Assign each customer to its nearest outlet



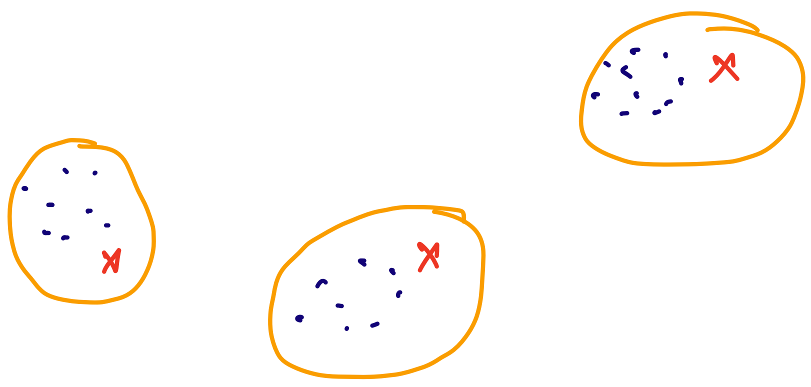
3. Update the center locations by taking the mean of

points assigned to the cluster.

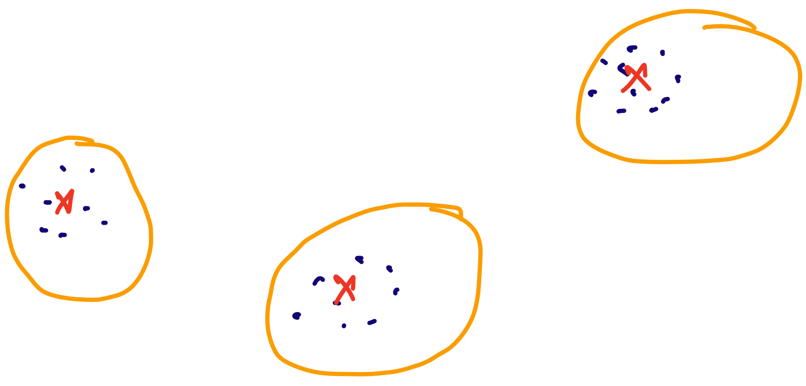


Step 2 & 3 repeat

Step 2:

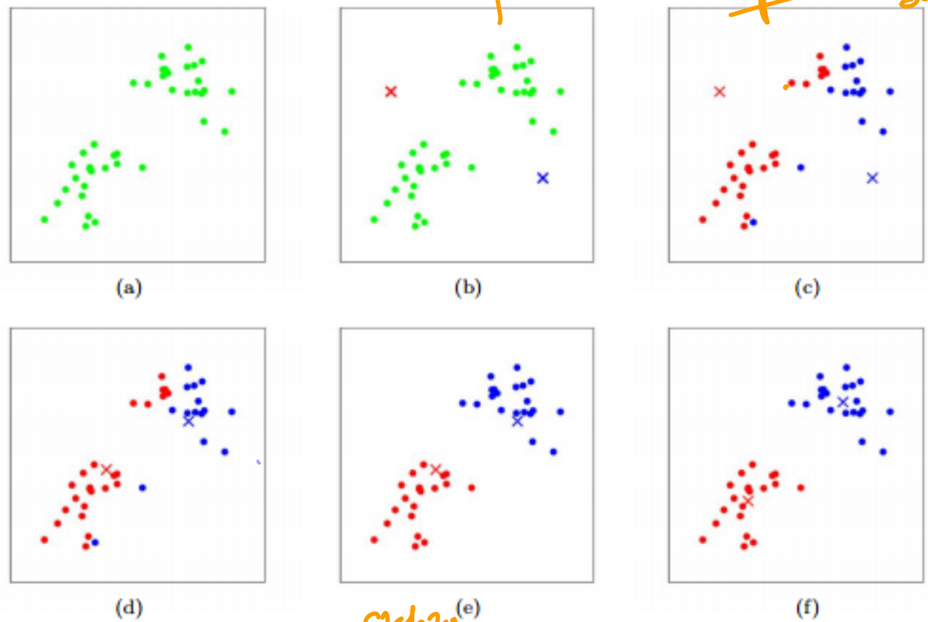


Step 3:



Randomly Initialise

Step2: Nearest centroid assign.

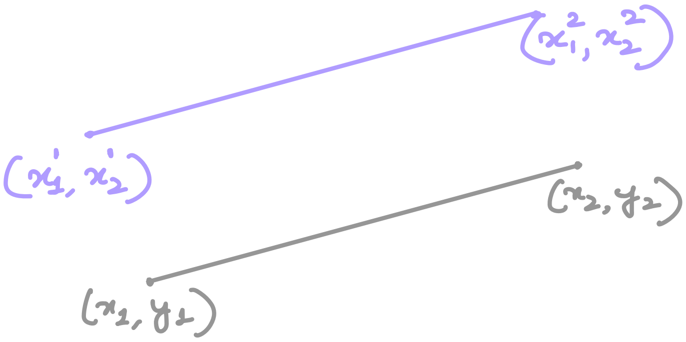


Step3: new mean

Step3: new mean

Step2: nearest centroid assign

Euclidean Distance



$$\sqrt{(x_2^2 - x_2^1)^2 + (x_1^2 - x_1^1)^2}$$

$$\sqrt{(y_2 - y_1)^2 + (x_2 - x_1)^2}$$