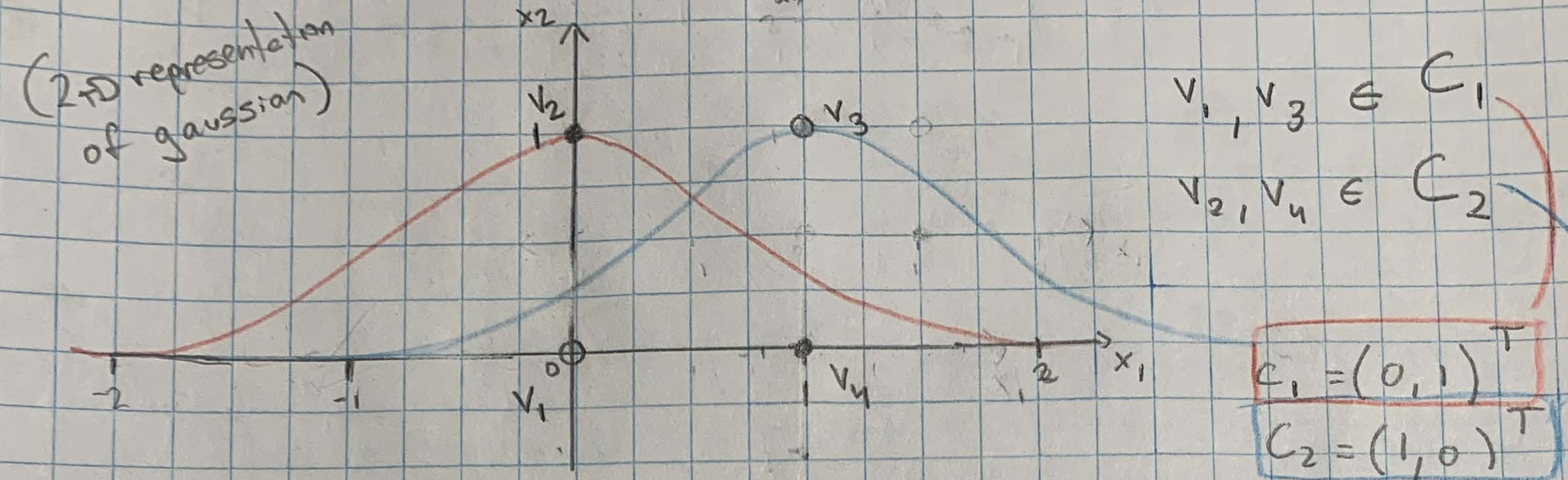
Cisc 452 A3 : Joseph Moraru 20181593

1)

XOR:

input vectors : $v_1(0,0)$, $v_2(0,1)$, $v_3(1,1)$, $v_4(1,0)$



$$\phi_n(x) = a\exp\left(-\frac{\|x - c_n\|^2}{2\sigma^2}\right)$$

c_n represents the center, a represents the vertical stretch and σ represents the horizontal stretch.

Chose
$a = 1$
$\sigma = \frac{1}{2}$

We can pick the first gaussian function for C_1, centered at $x = 0$ as :

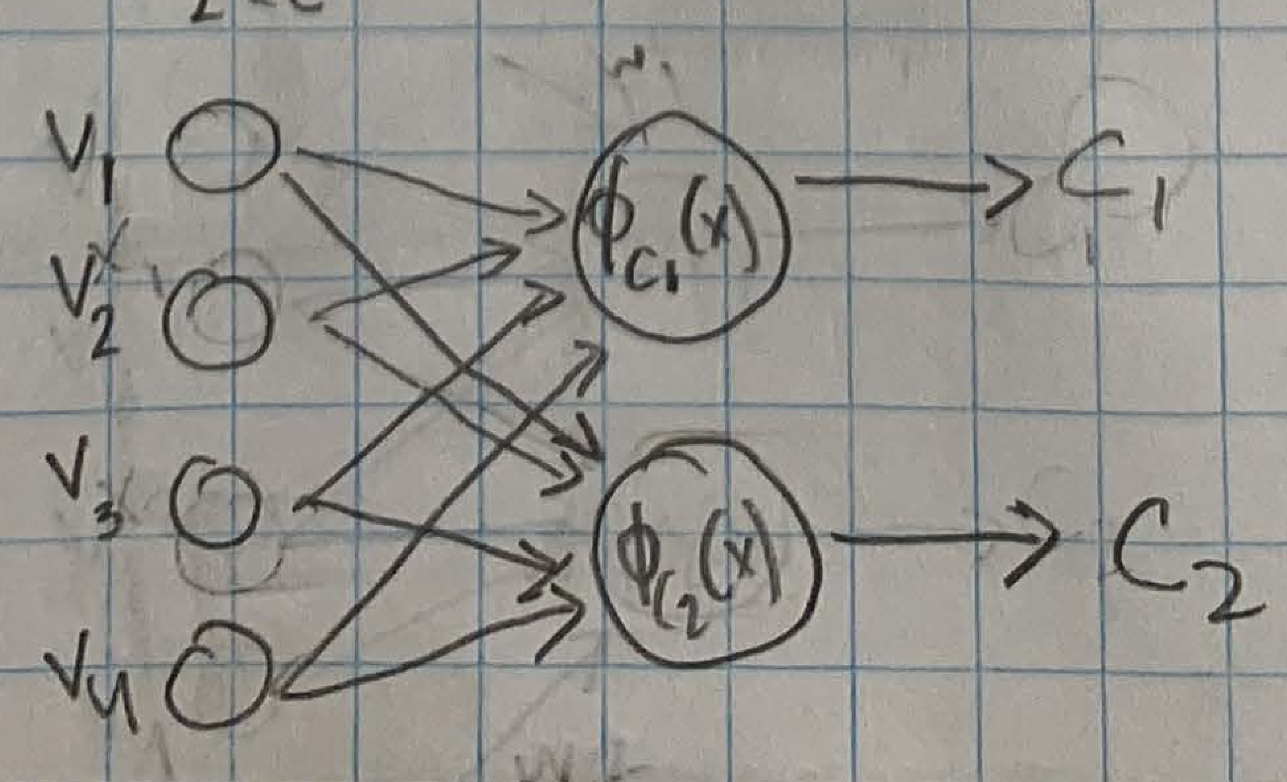
$$\phi_{c_1}(x) = e^{\left(-\frac{\|x-0\|^2}{2(\frac{1}{2})^2}\right)} = e^{\left(-\frac{\|x\|^2}{1/2}\right)}$$

We can pick the second gaussian function for C_2 centered at $x = 1$ as :

$$\phi_{c_2}(x) = e^{\left(-\frac{\|x-1\|^2}{2(\frac{1}{2})^2}\right)} = e^{\left(-\frac{\|x-1\|^2}{1/2}\right)}$$

	x_1	x_2	$\phi_{c_1}(x)$	$\phi_{c_2}(x)$	$\sum_{j=1}^{m} w_j A_j$, $A = \phi_j(x)$	Actual	Expected
v_1	0	0	1	e^2	e^2+1	C_1	C_1
v_2	0	1	1	1	$e^{2}+1$	C_2	C_2
v_3	1	1	e^{-2}	1	e^2+1	C_1	C_1
v_4	1	0	e^{-2}	e^2	$e^{-2}+e$	C_2	C_2

$w_1 = e^{2}$ $w_2 = e^{-2}$



Joseph Moraru 20181593

Cisc 452 : Assignment 3 Theoretical

2. $y(i) = \sum_{j=1}^{k} w_j(n) \exp\left(-\frac{1}{2\sigma^2(n)} \|\vec{x}(i) - \vec{\mu}_j(n)\|^2\right)$, $E = \frac{1}{2}\sum_{i=1}^{n} e^2(i)$, $e(i) = d(i) - y(i)$

a). Partial derivative w.r.t. $w_j(n)$:

$$\frac{\partial E}{\partial w_j(n)} = \frac{\partial E}{\partial y(i)} \cdot \frac{\partial y(i)}{\partial w_j(n)} = \frac{\partial}{\partial y(i)}\left(\frac{1}{2}\sum_{i=1}^{n}(d(i)-y(i))^2\right) \cdot \frac{\partial}{\partial w_j(n)}\left(\sum_{j=1}^{k} w_j(n)\exp\left(-\frac{1}{2\sigma^2(n)}\|\vec{x}(i)-\vec{\mu}_j(n)\|^2\right)\right)$$

$$\frac{\partial E}{\partial w_j(n)} = \frac{1}{2}\cdot(2)\cdot(d(i)-y(i))\left(\exp\left(-\frac{1}{2\sigma^2(n)}\|\vec{x}(i)-\vec{\mu}_j(n)\|^2\right)\right)$$

$$\boxed{\frac{\partial E}{\partial w_j(n)} = -(d(i)-y(i))\left(\exp\left(-\frac{1}{2\sigma^2(n)}\|\vec{x}(i)-\vec{\mu}_j(n)\|^2\right)\right)}$$

Partial Derivative w.r.t. $\vec{\mu}_j(n)$:

$$\frac{\partial E}{\partial \vec{\mu}_j(n)} = \frac{\partial E}{\partial y(i)} \cdot \frac{\partial (y(i))}{\partial \vec{\mu}_j(n)} = \frac{\partial}{\partial y(i)}\left(\frac{1}{2}\sum_{i=1}^{n}(d(i)-y(i))^2\right)\frac{\partial}{\partial \vec{\mu}_j(n)}\left(\sum_{j=1}^{k} w_j(n)\exp\left(-\frac{1}{2\sigma^2(n)}\|\vec{x}(i)-\vec{\mu}_j(n)\|^2\right)\right)$$

$$\frac{\partial E}{\partial \vec{\mu}_j(n)} = -(d(i)-y(i))\cdot\left(w_j(n)\exp\left(-\frac{1}{2\sigma^2(n)}\|\vec{x}(i)-\vec{\mu}_j(n)\|^2\right)\right)\frac{\partial}{\partial \vec{\mu}_j(n)}\left(-\frac{1}{2\sigma^2_n}\left((\vec{x}_1-\mu(1))^2+(\vec{x}_2-\mu(2))^2+(x_3-\mu(3))^2+\dots\right)\right)$$

$$= -(d(i)-y(i))\left(w_j(n)\exp\left(-\frac{1}{2\sigma^2(n)}\|\vec{x}(i)-\vec{\mu}_j(n)\|^2\right)\right)\left(-\frac{1}{2\sigma^2_n}\left(-2\,\vec{x}(i)-\vec{\mu}_j(n)\right)\right)$$

$$\boxed{\frac{\partial E}{\partial \vec{\mu}_j(n)} = -d(i)-y(i)\left(w_j(n)\exp\left(-\frac{1}{2\sigma^2(n)}\|\vec{x}(i)-\vec{\mu}_j(n)\|^2\right)\right)\left(\frac{1}{\sigma^2(n)}\left(\vec{x}(i)-\vec{\mu}_j(n)\right)\right)}$$

Partial Derivative w.r.t. $\sigma(n)$:

$$\frac{\partial E}{\partial \sigma(n)} = \frac{\partial E}{\partial y(i)} \cdot \frac{\partial (y(i))}{\partial \sigma(n)} = -(d(i)-y(i))\cdot\frac{\partial}{\partial \sigma(n)}\left(\sum_{j=1}^{k} w_j(n)\exp\left(-\frac{1}{2\sigma^2(n)}\|\vec{x}(i)-\vec{\mu}_j(n)\|^2\right)\right)$$

$$= -(d(i)-y(i))\left(\sum_{j=1}^{k} w_j(n)\exp\left(-\frac{1}{2\sigma^2(n)}\|\vec{x}(i)-\vec{\mu}_j(n)\|^2\right)\right)\left(\|\vec{x}_j(i)-\vec{\mu}_j(n)\|^2\right)\cdot\left(-\frac{1}{2}\right)(-2)\frac{1}{\sigma^3(n)}$$

$$\boxed{\frac{\partial E}{\partial \sigma(n)} = -(d(i)-y(i))\left(\sum_{j=1}^{k} w_j(n)\exp\left(-\frac{1}{2\sigma^2(n)}\|\vec{x}(i)-\vec{\mu}_j(n)\|^2\right)\right)\left(\frac{1}{\sigma^3(n)}\|\vec{x}_j(i)-\vec{\mu}_j(n)\|^2\right)}$$

See answer
from part a
(just substitute in here)

2b) Update formulas:

For $w_j(n) = w_j(n-1) - \eta_w \cdot \frac{\partial E}{\partial w_j(n)}$

For " $\mu_j(n) = \mu_j(n-1) - \eta_\mu \cdot \frac{\partial E}{\partial \mu_j(n)}$

For $\sigma(n) = \sigma(n-1) - \eta_\sigma \cdot \frac{\partial E}{\partial \sigma(n)}$

2c) The gradient vector, $\frac{\partial E}{\partial \vec{\mu}_j(n)}$, moves the center of the gaussian function by a certain magnitude, this is mainly done by the $\|\vec{x}_j(n) - \vec{\mu}_j(n)\|^2$ component of each partial derivative of the gradient vector.

When examining the result for $\frac{\partial E}{\partial \vec{\mu}_j(n)}$, specifically the

$$\frac{\partial E}{\partial \vec{\mu}_j(n)} = -(d(i) - y(i))\left(\sum_{j=1}^{k} w_j(n) \exp\left(-\frac{1}{2\sigma^2} \|x(i) - \mu(i)\|^2\right)\left(\frac{1}{\sigma^3(n)} \|\vec{x}(i) - \vec{\mu}(i)\|^2\right)\right)$$

this portion of $\frac{\partial E}{\partial \vec{\mu}_j(n)}$ very closely resembles clustering with its use of euclidean distance $\left(d(r_j, i_k) = \sqrt{\sum_{i=1}^{n} (i_{k,i} - w_{j,i})^2}\right)$ for calculating the closest centroids to a certain datapoint, and moving the centroids closer to those data points.

The weight update formula for many clustering algorithms closely resemble this gradient, $\frac{\partial E}{\partial \vec{\mu}_j(n)}$, in structure as well.

CISC 452 Assignment 3 Theoretical Joseph Moran 20181593

3) $x_1 = [0,0]^t$, $x_2 = [1,1]^t$, $x_3 = [-1,-1]^t$, $x_4 = [-2,2]^t$, $x_5 = [2,-2]^t$

$$X = \begin{bmatrix} 0 & 0 \\ 1 & 1 \\ -1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix}$$

i) $m_x = \frac{1}{n} \sum_{i=1}^{5} \vec{x}_i^{\,t}$ ← use this for calc. mean, m_x where $n = 5$

$m_x = \frac{1}{5}([0,0] + [1,1] + [-1,-1] + [-2,2] + [2,-2])$

$$\boxed{m_x = \frac{1}{5}([0,0]) = [0,0]}$$

ii) $C_x = \begin{bmatrix} cov(c_1,c_1) & cov(c_1,c_2) \\ cov(c_2,c_1) & cov(c_2,c_2) \end{bmatrix}$, where c_1 = column 1 of X, c_2 = column 2 of X

$cov(x_i, x_i) = \frac{1}{n} \sum_{i=1}^{x} [(x_i - m_x)(x_i - m_x)^T]$

$n \cdot cov(c_1,c_1) = (0-0)(0-0) + (1-0)(1-0) + (-1-0)(-1-0) + (-2-0)(-2-0) + (2-0)(2-0)$

$n \cdot cov(c_1,c_1) = 0 + 1 + 1 + 4 + 4 = 10$

$n \cdot cov(c_1,c_2) = (0-0)(0-0) + (1-0)(1-0) + (-1-0)(-1-0) + (-2-0)(2-0) + (2-0)(-2-0)$

$n \cdot cov(c_1,c_2) = 0 + 1 + 1 - 4 - 4 = -6$

$n \cdot cov(c_2,c_1) = 0 + 1 + 1 - 4 - 4 = -6$

$n \cdot cov(c_2,c_2) = 0 + 1 + 1 + 4 + 4 = 10$

$$\boxed{C_x = \frac{1}{5}\begin{bmatrix} 10 & -6 \\ -6 & 10 \end{bmatrix} = \begin{bmatrix} 2 & -\frac{6}{5} \\ -\frac{6}{5} & 2 \end{bmatrix}}$$

Cisc 452 Assignment 3 Joseph Moraru 20181593

3 ii) Recall that λ represents the eigenvalues
To find the eigenvectors we must calculate $\det(C_x - \lambda) = 0$

$$C_x - \lambda = \begin{bmatrix} 2 & -6/5 \\ -6/5 & 2 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$$

$$C_x - \lambda = \begin{bmatrix} 2-\lambda & -6/5 \\ -6/5 & 2-\lambda \end{bmatrix}$$

$$\det(C_x - \lambda) = (2-\lambda)(2-\lambda) - (-6/5)(-6/5) = 0$$

$$\det(C_x - \lambda) = 4 - 4\lambda + \lambda^2 - \frac{36}{25} = 0$$

$$0 = \frac{64}{25} - 4\lambda + \lambda^2$$

$$\lambda = \frac{4 \pm \sqrt{16 - 10.24}}{2} = \frac{16}{5} \text{ or } \frac{4}{5}$$

To find eigenvectors, we must slove $(C_x - \lambda)\sigma(C_x) = 0$ for each eigenvalue, λ, where $\sigma(C_x)$ represents the eigenvector.

Recall that $C_x - \lambda = \begin{bmatrix} 2-\lambda & -6/5 \\ -6/5 & 2-\lambda \end{bmatrix}$

$\boxed{\lambda = \frac{16}{5}}$

$$\begin{bmatrix} 2-\lambda & -6/5 \\ -6/5 & 2-\lambda \end{bmatrix} \sigma_1(C_x) = 0$$

$$\begin{bmatrix} -6/5 & -6/5 \\ -6/5 & -6/5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0$$

$$-6.5x - 6.5y = 0$$
$$-6.5x - 6.5y = 0$$

$$-x = y$$
$$x = -y$$

$$\boxed{\sigma_1(C_x) = \begin{bmatrix} 1 \\ -1 \end{bmatrix}}$$

$\boxed{\lambda = \frac{4}{5}}$

$$\begin{bmatrix} 2-\lambda & -6/5 \\ -6/5 & 2-\lambda \end{bmatrix} \sigma_2(C_x) = 0$$

$$\begin{bmatrix} 6/5 & -6/5 \\ -6/5 & 6/5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0$$

$$6/5x - 6/5y = 0$$
$$-6/5x + 6/5y = 0$$

$$x = y$$

$$\boxed{\sigma_2(C_x) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}}$$

$\therefore$, the eigenvectors and eigenvalues are $\lambda_1 = \frac{16}{5}$, $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and $\lambda_2 = \frac{4}{5}$, $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Cisc 452 : A3 : Joseph Moraru 20181593

4) Proof by Induction

(Base Case)

Suppose w_j^1 is the winning neuron, such that $n_j = 1$ and

$$w_j^{final} = w_j^1 = \frac{1}{n_j^1}\sum_{m=1}^{1} x^m = \frac{1}{1}\cdot x^m$$

We can see that from our presented step for setting the weight vectors (i.e. $w_j^1 = x^m$, $n_j = 1$), the theorem, $w_j^{final} = \frac{1}{n_j}\sum_{m=1}^{n_j} x^m$ works for the base case of $w_j^{final} = w_j^1$.

(Induction Hypothesis)

Now suppose the winning node is at iteration 'k' which exists within the sequence of clusters, so $k \in \{1, 2, \dots, C\}$ where C is the number of clusters in the clustering network.

By applying the theorem: $w_j^{final} = \frac{1}{n_j}\sum_{m=1}^{n_j} x^m$, we receive:

$$w_k^{final} = \frac{1}{k}\sum_{m=1}^{k} x^m,$$

we must substitute the other weight update formula $w_j^{new} = w_j^{old} + \frac{1}{n_j}(x^p - w_j^{old})$ and assume it is true for w_k^{final} such that

$$w_k^{final} = \frac{1}{k}\sum_{m=1}^{k} x^m = w_{k-1} + \frac{1}{k}(x^k - w_{k-1})$$

(Induction step) Assuming the above, if we can show that the formula works for winning node $k+1 \in \{1, 2, \dots, C\}$ then we can prove the theorem works for any $k \in \{1, 2, \dots, C\}$.

$$w_{k+1}^{final} = \frac{1}{k+1}\sum_{m=1}^{k+1} x^m$$

$$w_{k+1} = w_k + \frac{1}{k+1}(x^{k+1} - w_k),$$

from the induction hypothesis we know that $w_k = \frac{1}{k}\sum_{m=1}^{k} x^m$, we can substitute this in.

$$w_{k+1} = \frac{1}{k}\sum_{m=1}^{k} x^k + \frac{1}{k+1}\left(x^{k+1} - \frac{1}{k}\sum_{m=1}^{k} x^k\right)$$

$$w_{k+1} = \frac{1}{k+1}(x^{k+1}) - \frac{1}{k}\cdot\frac{1}{k+1}\sum_{m=1}^{k} x^k + \frac{1}{k}\sum_{m=1}^{k} x^k$$

$$w_{k+1} = \frac{1}{k+1}(x^{k+1}) + \frac{1}{k}\sum_{m=1}^{k} x^k\left(1 - \frac{1}{k+1}\right)$$

$$\boxed{w_{k+1} = \frac{1}{k+1}\sum_{m=1}^{k+1} x^{k+1} = \frac{1}{k+1}\sum_{m=1}^{k+1} x^m}$$

∴, by proof of induction we have shown that after presentation of all training sets, the theorem holds true.