Homework 1, Data Mining, Fall 2014

Aaron Gonzales

September 11, 2014

given the following vectors:

$$x = \{1, 3, 5, 7, 6, 8, 5, 2\} \tag{1}$$

$$y = \{7, 11, 9, 6, 4, 3, 1, 3\} \tag{2}$$

$$z = \{3, 5, 7, 9, 8, 7, 6, 5\} \tag{3}$$

1 wavelet transform each vector

An un-normalized 8-point Haar matrix

 H_8

is shown below

and the vector

$$\{\sqrt{8}, \sqrt{8}, 2, 2, \sqrt{2}, \sqrt{2}, \sqrt{2}, \sqrt{2}\}$$

and will be used to normalize each vector after transformation.

(a) vector x

$$\{37, -4, -8, 7, -2, -2, -2, 3\}$$

$$\{\sqrt{8}, \sqrt{8}, 2, 2, \sqrt{2}, \sqrt{2}, \sqrt{2}, \sqrt{2}, \sqrt{2}, \}$$

$$= \{74\sqrt{2}, -8\sqrt{2}, -16, 14, -2\sqrt{2}, -2\sqrt{2}, -2\sqrt{2}, 3\sqrt{2}\}$$

$$\{104.652, -11.3137, -16., 14., -2.82843, -2.82843, 4.24264\}$$

(b) vector y

$$= \{44,22,3,3,-4,3,1,-2\}$$

$$\{\sqrt{8},\sqrt{8},2,2,\sqrt{2},\sqrt{2},\sqrt{2},\sqrt{2},\sqrt{2},\}$$

$$= \{88\sqrt{2},44\sqrt{2},6,6,-4\sqrt{2},3\sqrt{2},\sqrt{2},-2\sqrt{2}\}$$

$$\{124.451,62.2254,6.,6.,-5.65685,4.24264,1.41421,-2.82843\}$$

(c) vector z

$$\{50, -2, -8, 4, -2, -2, 1, 1\}$$

$$\{\sqrt{8}, \sqrt{8}, 2, 2, \sqrt{2}, \sqrt{2}, \sqrt{2}, \sqrt{2}, \sqrt{2}, \}$$

$$= \{100\sqrt{2}, -4\sqrt{2}, -16, 8, -2\sqrt{2}, -2\sqrt{2}, \sqrt{2}, \sqrt{2}\}$$

$$\{141.421, -5.65685, -16, 8, -2.82843, 1.41421, 1.41421\}$$

2 fourier transform each vector

(Defintion from Wikipedia)

The transformation W of size $N \times N$ can be defined as

$$W = \left(\frac{\omega^{jk}}{\sqrt{N}}\right)_{j,k=0,\dots,N-1}$$

, or equivalently:

$$W = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & 1 & 1 & 1 & \cdots & 1\\ 1 & \omega & \omega^2 & \omega^3 & \cdots & \omega^{N-1}\\ 1 & \omega^2 & \omega^4 & \omega^6 & \cdots & \omega^{2(N-1)}\\ 1 & \omega^3 & \omega^6 & \omega^9 & \cdots & \omega^{3(N-1)}\\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & \omega^{N-1} & \omega^{2(N-1)} & \omega^{3(N-1)} & \cdots & \omega^{(N-1)(N-1)} \end{pmatrix}$$

where $\omega = e^{-\frac{2\pi i}{N}}$

In our case, with an 8×8 matrix, we get a transformation matrix of:

$$W = \frac{1}{\sqrt{8}} \begin{pmatrix} \omega^{0} & \omega^{0} & \omega^{0} & \dots & \omega^{0} \\ \omega^{0} & \omega^{1} & \omega^{2} & \dots & \omega^{7} \\ \omega^{0} & \omega^{2} & \omega^{4} & \dots & \omega^{14} \\ \omega^{0} & \omega^{3} & \omega^{6} & \dots & \omega^{21} \\ \omega^{0} & \omega^{4} & \omega^{8} & \dots & \omega^{28} \\ \omega^{0} & \omega^{5} & \omega^{10} & \dots & \omega^{35} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \omega^{0} & \omega^{7} & \omega^{14} & \dots & \omega^{49} \end{pmatrix}$$

where $\omega=e^{-\frac{2\pi i}{8}}=\frac{1}{\sqrt{2}}-\frac{i}{\sqrt{2}}$ where $\omega=e^{-2\pi i}=1$ is Euler's identity, allowing us to reduce the 8×8 matrix as follows:

$$\frac{1}{\sqrt{8}} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & e^{\frac{-\pi i}{4}} & -i & e^{\frac{-3\pi i}{4}} & -1 & e^{\frac{-5\pi i}{4}} \\ 1 & -i & -1 & i & 1 & -i \\ 1 & e^{\frac{-3\pi i}{4}} & i & e^{\frac{-\pi i}{4}} & -1 & e^{\frac{-7\pi i}{4}} \\ 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & e^{\frac{-5\pi i}{4}} & -i & e^{\frac{-7\pi i}{4}} & -1 & e^{\frac{-\pi i}{4}} \\ 1 & i & -1 & -i & 1 & i \\ 1 & e^{\frac{-7\pi i}{4}} & i & e^{\frac{-5\pi i}{4}} & -1 & e^{\frac{-3\pi i}{4}} \end{pmatrix}$$

(a) Vector x

$$x = \{1, 3, 5, 7, 6, 8, 5, 2\}$$

$$\frac{1}{\sqrt{8}} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & e^{\frac{-\pi i}{4}} & -i & e^{\frac{-3\pi i}{4}} & -1 & e^{\frac{-5\pi i}{4}} \\ 1 & -i & -1 & i & 1 & -i \\ 1 & e^{\frac{-3\pi i}{4}} & i & e^{\frac{-\pi i}{4}} & -1 & e^{\frac{-7\pi i}{4}} \\ 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & e^{\frac{-3\pi i}{4}} & -i & e^{\frac{-\pi i}{4}} & -1 & e^{\frac{-\pi i}{4}} \\ 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & e^{\frac{-5\pi i}{4}} & -i & e^{\frac{-7\pi i}{4}} & -1 & e^{\frac{-\pi i}{4}} \\ 1 & i & -1 & -i & 1 & i \\ 1 & e^{\frac{-7\pi i}{4}} & i & e^{\frac{-5\pi i}{4}} & -1 & e^{\frac{-3\pi i}{4}} \\ \end{pmatrix} * \begin{pmatrix} 1 \\ 3 \\ 5 \\ 7 \\ 6 \\ 8 \\ 5 \\ 2 \end{pmatrix} * \begin{pmatrix} 1 \\ 3 \\ 5 \\ 7 \\ 6 \\ 8 \\ 5 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 + 3 + 5 + 7 + 6 + 8 + 5 + 2 \\ 1 + 3 e^{\frac{-\pi i}{4}} + 5 + 7 e^{\frac{-3\pi i}{4}} - 6 + 8 e^{\frac{-5\pi i}{4}} + 5 i + 2 e^{\frac{-7\pi i}{4}} \\ 1 - 3 i + 5 + 7 i + 6 - 8 i - 5 + 2 i \\ 1 + 3 e^{\frac{-3\pi i}{4}} + 5 i + 7 e^{\frac{-\pi i}{4}} - 6 + 8 e^{\frac{-7\pi i}{4}} - 5 i + 2 e^{\frac{-5\pi i}{4}} \\ 1 - 3 + 5 - 7 + 6 - 8 + 5 - 2 \\ 1 + 3 e^{\frac{-5\pi i}{4}} - 5 i + 7 e^{\frac{-7\pi i}{4}} - 6 + 8 e^{\frac{-\pi i}{4}} + 5 + 2 e^{\frac{-3\pi i}{4}} \\ 1 + 3 i - 5 - 7 i + 6 + 8 i - 5 - 2 i \\ 1 + 3 e^{\frac{-7\pi i}{4}} + 5 i + 7 e^{\frac{-5\pi i}{4}} - 6 + 8 e^{\frac{-3\pi i}{4}} - 5 i + 2 e^{\frac{-3\pi i}{4}} \end{pmatrix}$$

$$= \begin{pmatrix} 37 \\ -5 + 3e^{-i\pi/4} + 2e^{i\pi/4} + 7e^{-3i\pi/4} + 8e^{3i\pi/4} \\ -3 - 2i \\ -5 + 7e^{-i\pi/4} + 8e^{i\pi/4} + 3e^{-3i\pi/4} + 2e^{3i\pi/4} \\ -3 \\ -5i + 8e^{-i\pi/4} + 7e^{i\pi/4} + 2e^{-3i\pi/4} + 3e^{3i\pi/4} \\ -3 + 2i \\ -5 + 2e^{-i\pi/4} + 3e^{i\pi/4} + 8e^{-3i\pi/4} + 7e^{3i\pi/4} \end{pmatrix} = \begin{pmatrix} 37 \\ -12.07106 \\ -3 - 2i \\ 2.07106 \\ -3 \\ 7.07106 - 5i \\ -3 + 2i \\ -12.07106 \end{pmatrix} * \frac{1}{\sqrt{8}}$$

$$\{13.0815, -4.26776, -1.06066 - 0.707107i, 0.73223, -1.06066, 2.5 - 1.76777i, -1.06066 + 0.707107i, -4.26776\}$$

(b) y vector

$$y = \{7,11,9,6,4,3,1,3\}$$

$$\frac{1}{\sqrt{8}} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & e^{\frac{-\pi i}{4}} & -i & e^{\frac{-3\pi i}{4}} & -1 & e^{\frac{-5\pi i}{4}} \\ 1 & -i & -1 & i & 1 & -i \\ 1 & e^{\frac{-3\pi i}{4}} & i & e^{\frac{-\pi i}{4}} & -1 & e^{\frac{-7\pi i}{4}} \\ 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & e^{\frac{-5\pi i}{4}} & -i & e^{\frac{-7\pi i}{4}} & -1 & e^{\frac{-\pi i}{4}} \\ 1 & i & -1 & -i & 1 & i \\ 1 & e^{\frac{-5\pi i}{4}} & -i & e^{\frac{-7\pi i}{4}} & -1 & e^{\frac{-\pi i}{4}} \\ 1 & i & -1 & -i & 1 & i \\ 1 & e^{\frac{-5\pi i}{4}} & i & e^{\frac{-5\pi i}{4}} & -1 & e^{\frac{-3\pi i}{4}} \end{pmatrix} * \begin{pmatrix} 7\\11\\9\\6\\4\\3\\1\\1\\3 \end{pmatrix} = \begin{pmatrix} 7+11+9+6+4+3+1+3\\7+11e^{\frac{-3\pi i}{4}}+9i+6e^{\frac{-3\pi i}{4}}+4+3e^{\frac{-5\pi i}{4}}+i+3e^{\frac{-7\pi i}{4}}\\7+11e^{\frac{-3\pi i}{4}}+9i+6e^{\frac{-\pi i}{4}}+4+3e^{\frac{-7\pi i}{4}}+i+3e^{\frac{-5\pi i}{4}}\\7+11e^{\frac{-5\pi i}{4}}+9i+6e^{\frac{-7\pi i}{4}}+-4+3e^{\frac{-\pi i}{4}}+1+3e^{\frac{-3\pi i}{4}}\\7+11e^{\frac{-7\pi i}{4}}+9i+6e^{\frac{-7\pi i}{4}}+-4+3e^{\frac{-3\pi i}{4}}+1+3e^{\frac{-3\pi i}{4}}\\7+11e^{\frac{-7\pi i}{4}}+9i+6e^{\frac{-5\pi i}{4}}+-4+3e^{\frac{-3\pi i}{4}}+1+3e^{\frac{-\pi i}{4}}\end{pmatrix}$$

$$\approx \{15.5563, 2.31066 - 5.57843i, 0.353553 - 1.76777i, \\ -0.189338 + 0.0784252i, -0.707107, 0.164211 - 0.43198i, \\ 0.353553 + 1.76777i, 2.31065 + 5.5784i\}$$

(c) z vector

$$z = \{3, 5, 7, 9, 8, 7, 6, 5\}$$

$$\frac{1}{\sqrt{8}}\begin{pmatrix}1&1&1&1&1&1\\1&e^{\frac{-\pi i}{4}}&-i&e^{\frac{-3\pi i}{4}}&-1&e^{\frac{-5\pi i}{4}}\\1&-i&-1&i&1&-i\\1&e^{\frac{-3\pi i}{4}}&i&e^{\frac{-\pi i}{4}}&-1&e^{\frac{-7\pi i}{4}}\\1&-1&1&-1&1&-1\\1&e^{\frac{-5\pi i}{4}}&-i&e^{\frac{-7\pi i}{4}}&-1&e^{\frac{-\pi i}{4}}\\1&i&-1&1&-1&1&-1\\1&e^{\frac{-5\pi i}{4}}&-i&e^{\frac{-7\pi i}{4}}&-1&e^{\frac{-\pi i}{4}}\\1&i&-1&-i&1&i\\1&e^{\frac{-7\pi i}{4}}&i&e^{\frac{-5\pi i}{4}}&-1&e^{\frac{-3\pi i}{4}}\\1&i&-1&-i&1&i\\1&e^{\frac{-7\pi i}{4}}&i&e^{\frac{-5\pi i}{4}}&-1&e^{\frac{-3\pi i}{4}}\\1&i&e^{\frac{-7\pi i}{4}}&i&e^{\frac{-5\pi i}{4}}&-1&e^{\frac{-3\pi i}{4}}\\1&e^{\frac{-7\pi i}{4}}&e^{\frac{-7\pi i}{4}}&-1&e^{\frac{-3\pi i}{4}}&-1&e^{\frac{-7\pi i}{4}}\\1&e^{\frac{-7\pi i}{4}}&e^{\frac{-7\pi i}{4}}&-1&e^{\frac{-3\pi i}{4}}&-1&e^{\frac{-3\pi i}{4}}\\1&e^{\frac{-7\pi i}{4}}&e^{\frac{-7\pi i}{4}}&-1&e^{\frac{-7\pi i}{4}}&-1&e$$

$$=\begin{pmatrix} 50 \\ (-5-i)+5e^{-(i\pi)/4}+5e^{i\pi)/4}+9e^{-(3i\pi)/4}+7e^{3i\pi)/4} \\ -7+7i \\ (-5+i)+5^{-(3i\pi)/4}+9e^{-(i\pi)/4}+7e^{(i\pi)/4}+5e^{(3i\pi)/4} \\ -2 \\ 3+5^{1/4(-5\pi i)}-7i+9e^{1/4(-7\pi i)}-8+7e^{-(\pi i)/4}+6+5e^{1/4(-3\pi i)} \\ 3-7i \\ (-5+i)+5^{-(7i\pi)/4}+5e^{-(i\pi)/4}+7e^{-(3i\pi)/4}+9e^{(3i\pi)/4} \end{pmatrix} = \begin{pmatrix} 50 \\ -9.242-2.4142i \\ -7+7i \\ 1.9824+3.7269i \\ -2 \\ 9.7774-9.1583i \\ 3-7i \\ -13.6165-1.6663i \end{pmatrix} *\frac{1}{\sqrt{8}}$$

$$\begin{aligned} \{17.6777, -3.26754 - 8535.49i, -2.47487 + 2.47487i, \\ 0.700884 + 1.31766i, -0.707107, 3.45683 - 3.23795i, \\ 1.06066 - 2.47487i, -4.81416 - 0.589126i\} \end{aligned}$$

3 minkowski distance

The Minkowski distance of order *p* between two points

$$P = (x_1, x_2, ..., x_n)$$
 and $Q = (y_1, y_2, ..., y_n) \in \mathbb{R}^n$

is defined as:

$$\left(\sum_{i=1}^{n}|x_i-y_i|^p\right)^{1/p}$$

(Wikipedia)

and can be written as:

$$d(i,j) = \sqrt[h]{|x_{i1} - x_{j1}|^h + |x_{i2} - x_{j2}|^h + \dots + |x_{ip} - x_{jp}|^h}$$

where $i = \{x_{i1}, x_{i_2}, \dots, x_{ip}\}$ and $j = \{x_{j1}, x_{j_2}, \dots, x_{jp}\}$ are two *p*-dimensional data objects and *h* is the order (from class slides).

(a) raw vectors

Between x and y:

$$x = \{1, 3, 5, 7, 6, 8, 5, 2\}$$

$$y = \{7, 11, 9, 6, 4, 3, 1, 3\}$$

$$\sqrt{|1-7|^2+|3-11|^2+|5-9|^2+|7-6|^2+|6-4|^2+|8-3|^2+|5-1|^2+|2-3|^2}$$

Aaron Gonzales; Data Mining

Homework 1

$$= \sqrt{163}$$

$$\approx 12.767$$

Between x and z:

$$x = \{1, 3, 5, 7, 6, 8, 5, 2\}$$

$$z = \{3, 5, 7, 9, 8, 7, 6, 5\}$$

$$\sqrt{|1-3|^2 + |3-5|^2 + |5-7|^2 + |7-9|^2 + |6-8|^2 + |8-7|^2 + |5-6|^2 + |2-5|^2}$$

$$= \sqrt{13}$$

$$\approx 5.56776$$

between y and z:

$$y = \{7, 11, 9, 6, 4, 3, 1, 3\}$$

$$z = \{3, 5, 7, 9, 8, 7, 6, 5\}$$

$$\sqrt{|7-3|^2 + |11-5|^2 + |9-7|^2 + |6-9|^2 + |4-8|^2 + |3-7|^2 + |1-6|^2 + |3-5|^2}$$

$$= 3\sqrt{14}$$

$$\approx 11.2249$$

$$\begin{pmatrix} 1 & 3 & 5 & 7 & 6 & 8 & 5 & 2 \\ 7 & 11 & 9 & 6 & 4 & 3 & 1 & 3 \\ 3 & 5 & 7 & 9 & 8 & 7 & 6 & 5 \end{pmatrix}$$

(b) wavlet coeff.

 $X: \{104.652, -11.3137, -16., 14., -2.82843, -2.82843, -2.82843, 4.24264\}$

Y: {124.451,62.2254,6.,6., -5.65685,4.24264,1.41421, -2.82843}

Z: {141.421, -5.65685, -16, 8, -2.82843, -2.82843, 1.41421, 1.41421} Between x and y:

$$\sqrt{|104.652 - 124.451|^2 + |-11.3137 - 62.2254|^2 + |-16 - 6|^2 + |14 - 6|^2 + |-2.82843 - -5.65685|^2 + |-2.82843 - 4|^2 + |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |-2.82843 - |$$

between y and z Y: {124.451,62.2254,6.,6., -5.65685,4.24264,1.41421, -2.82843} Z: {141.421, -5.65685, -16,8, -2.82843, -2.82843, 1.41421, 1.41421}

$$\sqrt{|124.451+141.451|^2+|62.2254--5.65685|^2+|6--16|^2} + \\ \overline{|6-8|^2+|-5.65685-2.82843|^2+|4.24264-2.82843|^2+|1.41421--2.82843-1.41421|^2}$$

$$\approx 275.485$$

X: {104.652, -11.3137, -16., 14., -2.82843, -2.82843, -2.82843, 4.24264} Z: {141.421, -5.65685, -16, 8, -2.82843, -2.82843, 1.41421, 1.41421}

$$\begin{array}{c} \sqrt{|104.652-141.421|^2+|-11.3137--5.65658|^2+|-16-16|^2+}\\ \hline \\ \overline{|14-8|^2+|-2.82843--2.82843|^2+|-2.82843--2.82843|^2+}\\ \hline \\ \overline{|-2.82843-1.41421|^2+|4.24264-1.41421|^2}\\ \approx 49.6987 \end{array}$$

- (c) largest three wavelet coefficients of x
- (d) largest three wavelet coefficients of y
- (e) which one of the above approximations matches the ordering of the distances from the raw vectors?