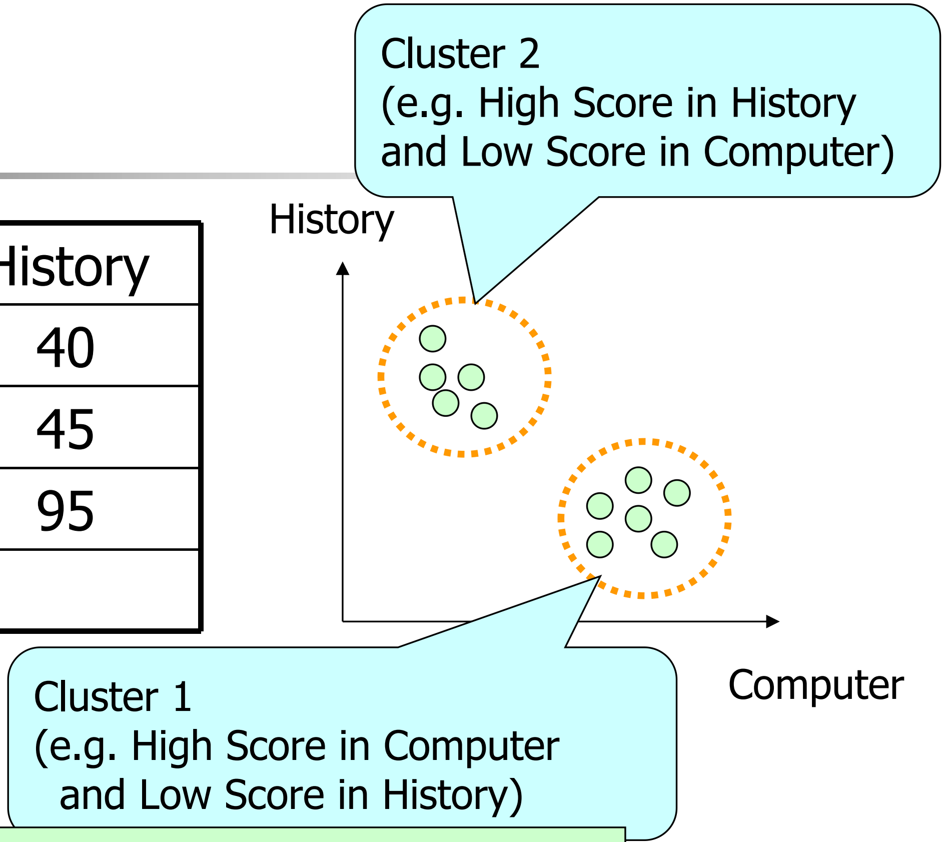
CSIT5210

Subspace Clustering

Prepared by Raymond Wong
Presented by Raymond Wong
raywong@cse

Clustering

	Computer	History
Raymond	100	40
Louis	90	45
Wyman	20	95
…	…	…

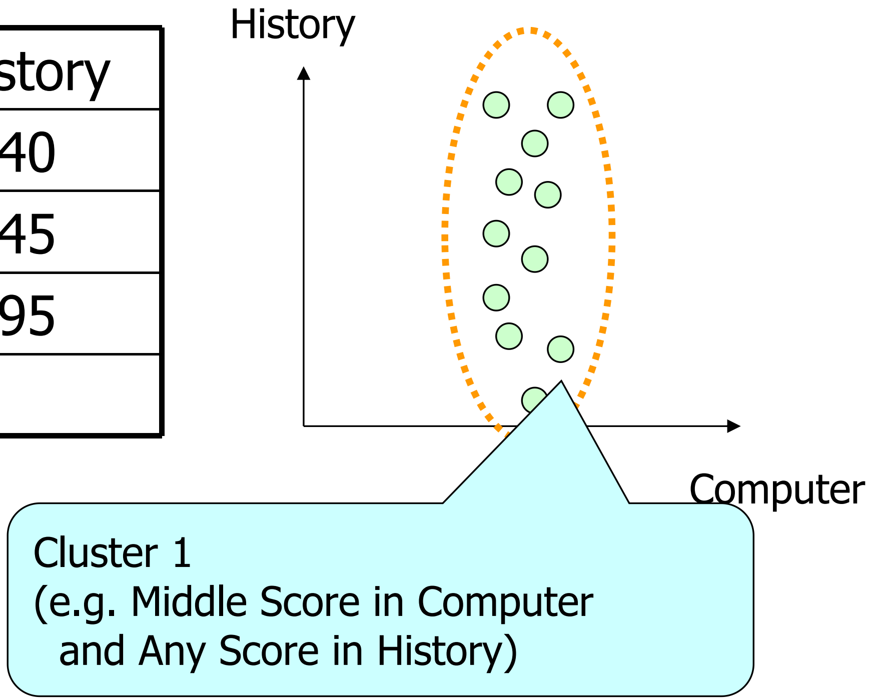


Problem: to find all clusters

This kind of clustering considers only FULL space (i.e. computer and history)!

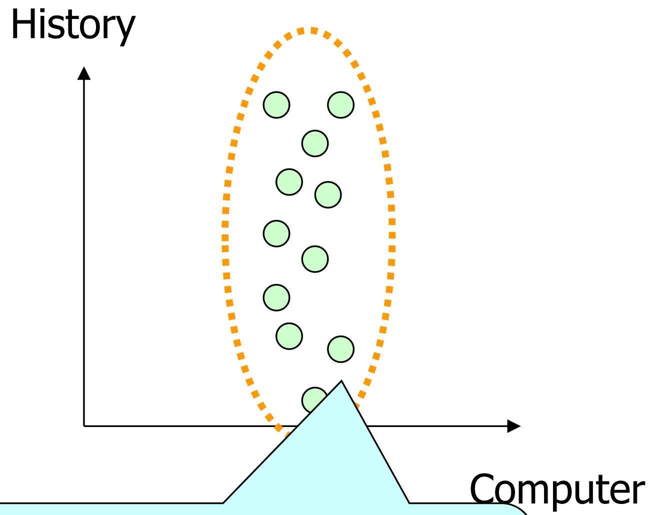
Subspace Clustering

	Computer	History
Raymond	50	40
Louis	60	45
Wyman	40	95
…	…	…



Subspace Clustering

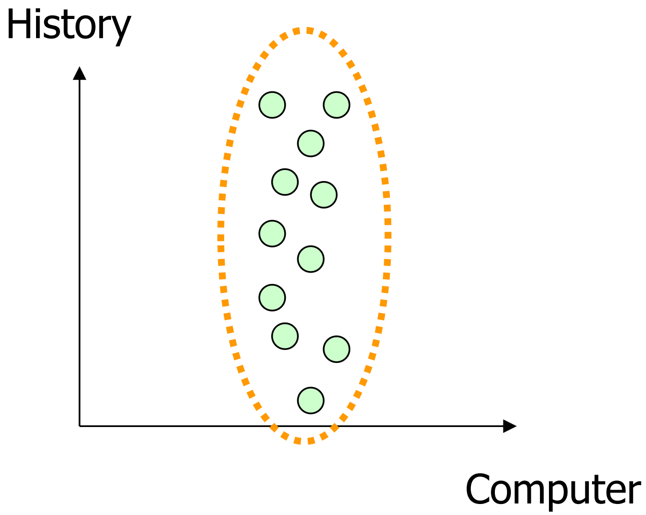
	Computer	History
Raymond	50	40
Louis	60	45
Wyman	40	95
…	…	…



Cluster 1
(e.g. Middle Score in Computer
~~and Any Score in History)~~

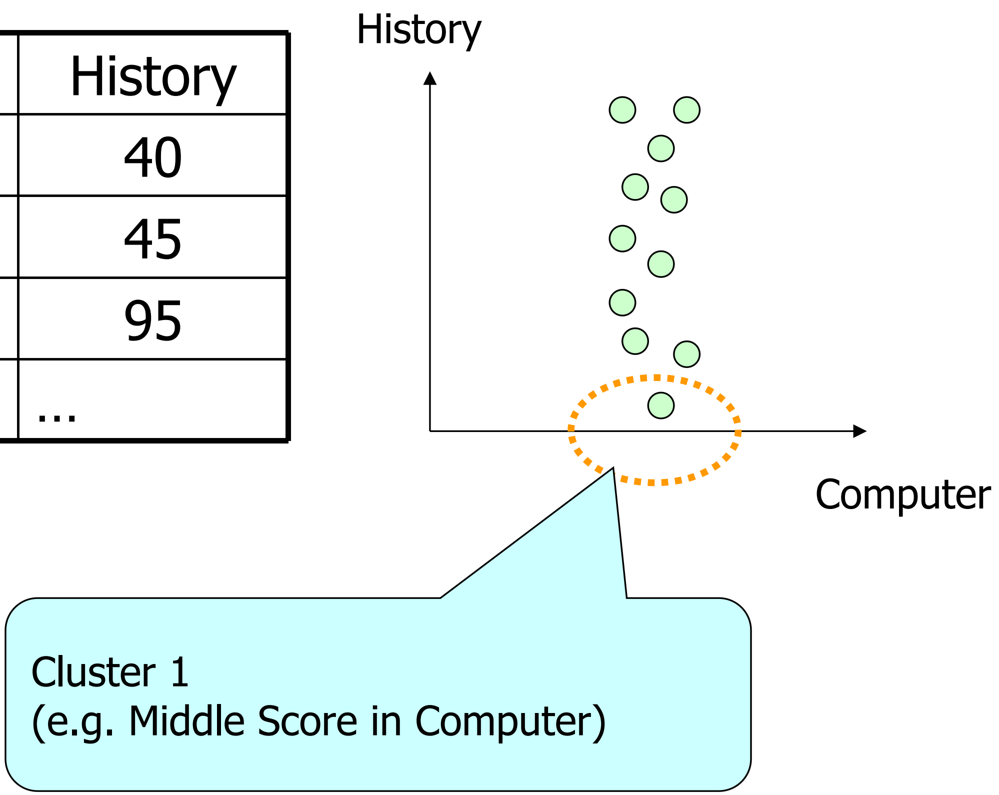
Subspace Clustering

	Computer	History
Raymond	50	40
Louis	60	45
Wyman	40	95
…	…	…



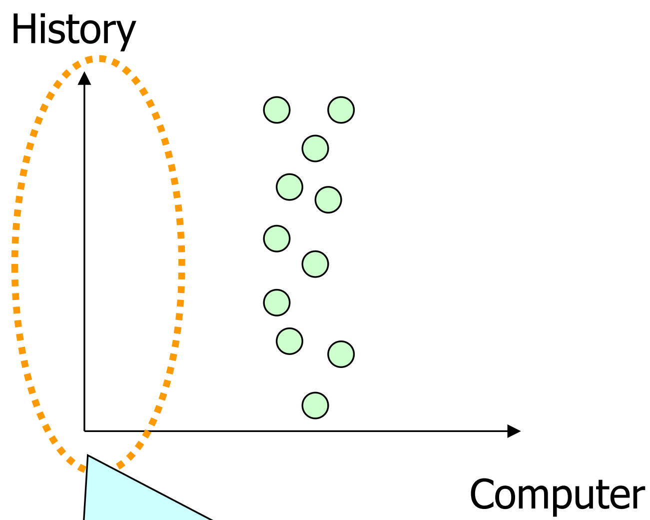
Subspace Clustering

	Computer	History
Raymond	50	40
Louis	60	45
Wyman	40	95
…	…	…



Subspace Clustering

	Computer	History
Raymond	50	40
Louis	60	45
Wyman	40	95
...	...	...



No Cluster!
The data points span along history dimension.

Problem: to find all clusters in the subspace (i.e. some of the dimensions)

Why Subspace Clustering?

- Clustering for Understanding
  - Applications
    - Biology
      - Group different species
    - Psychology and Medicine
      - Group medicine
    - Business
      - Group different customers for marketing
    - Network
      - Group different types of traffic patterns
    - Software
      - Group different programs for data analysis

Curse of Dimensionality

- When the number of dimensions increases,
  - the distance between any two points is nearly the same

Surprising results!

This is the reason why we need to study subspace clustering

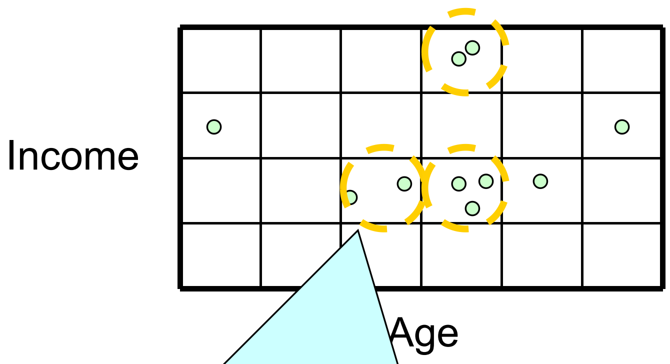
Subspace Clustering Methods

- Dense Unit-based Method
- Entropy-Based Method
- Transformation-Based Method

Dense Unit-based Method for Subspace Clustering

Density

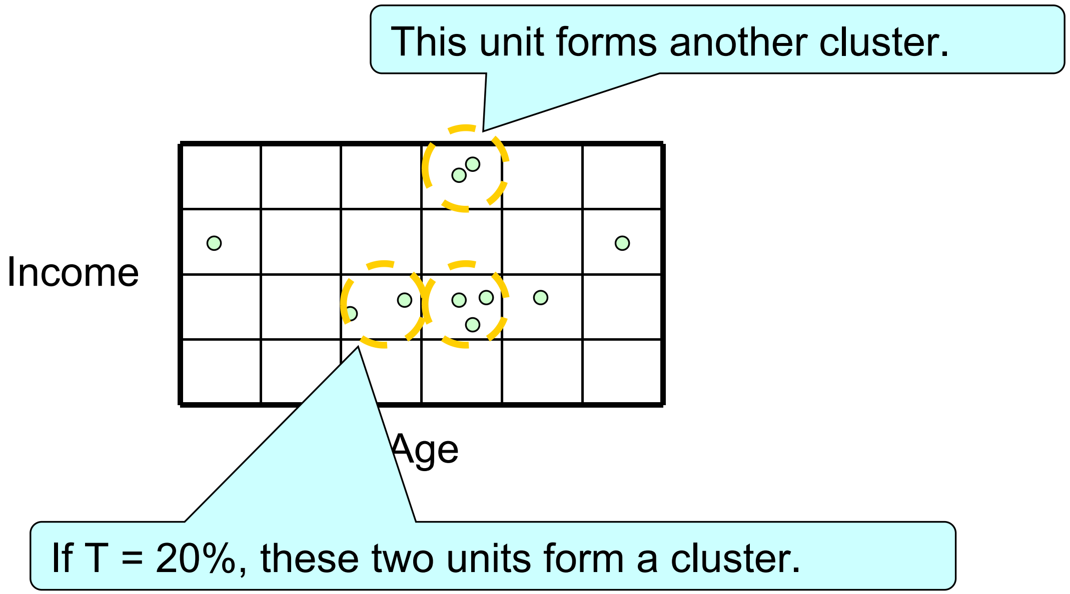
Dense unit: a unit if the fraction of data points contained in it is at least a threshold, T



If T = 20%, these three units are dense.

Dense Unit-based Method for Subspace Clustering

Cluster: a maximal set of connected dense units in k-dimensions



The problem is to find which **sub-spaces** contain dense units.
The second problem is to find clusters from each sub-space containing dense units

Dense Unit-based Method for Subspace Clustering

- **Step 1:** Identify sub-spaces that contain dense units
- **Step 2:** Identify clusters in each sub-spaces that contain dense units

Step 1

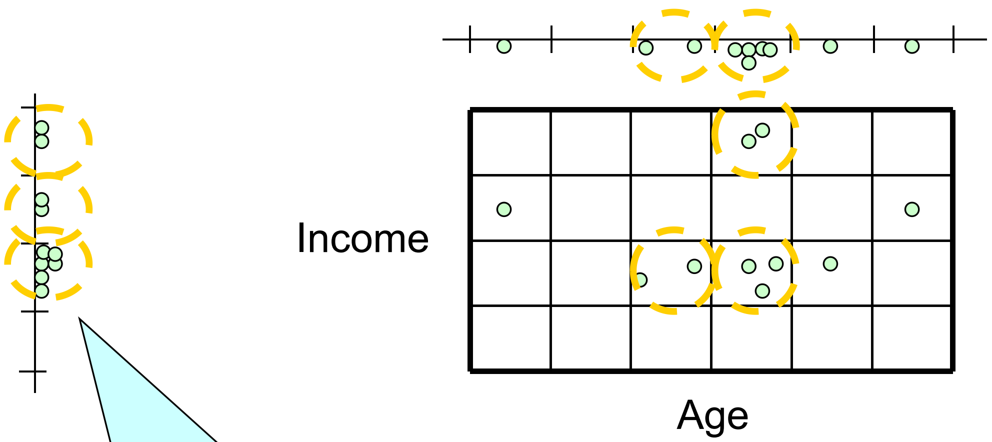
Suppose we want to find all dense units (e.g., dense units with density >= 20%)

- **Property**
  - If a set S of points is a cluster in a k-dimensional space, then S is also part of a cluster in any (k-1)-dimensional projections of the space.

Step 1

Suppose we want to find all dense units (e.g., dense units with density >= 20%)

If T = 20%, these two units are dense.



If T = 20%, these three units are dense.

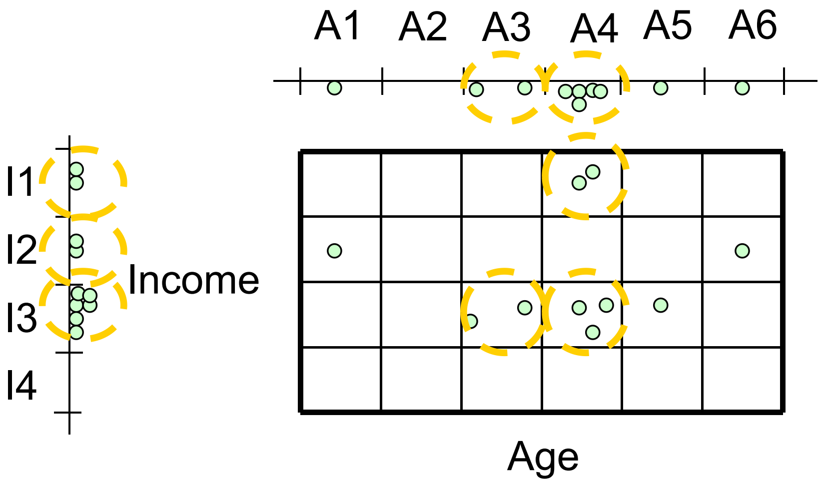
Step 1

Suppose we want to find all dense units (e.g., dense units with density >= 20%)

- We can make use of apriori approach to solve the problem

Step 1

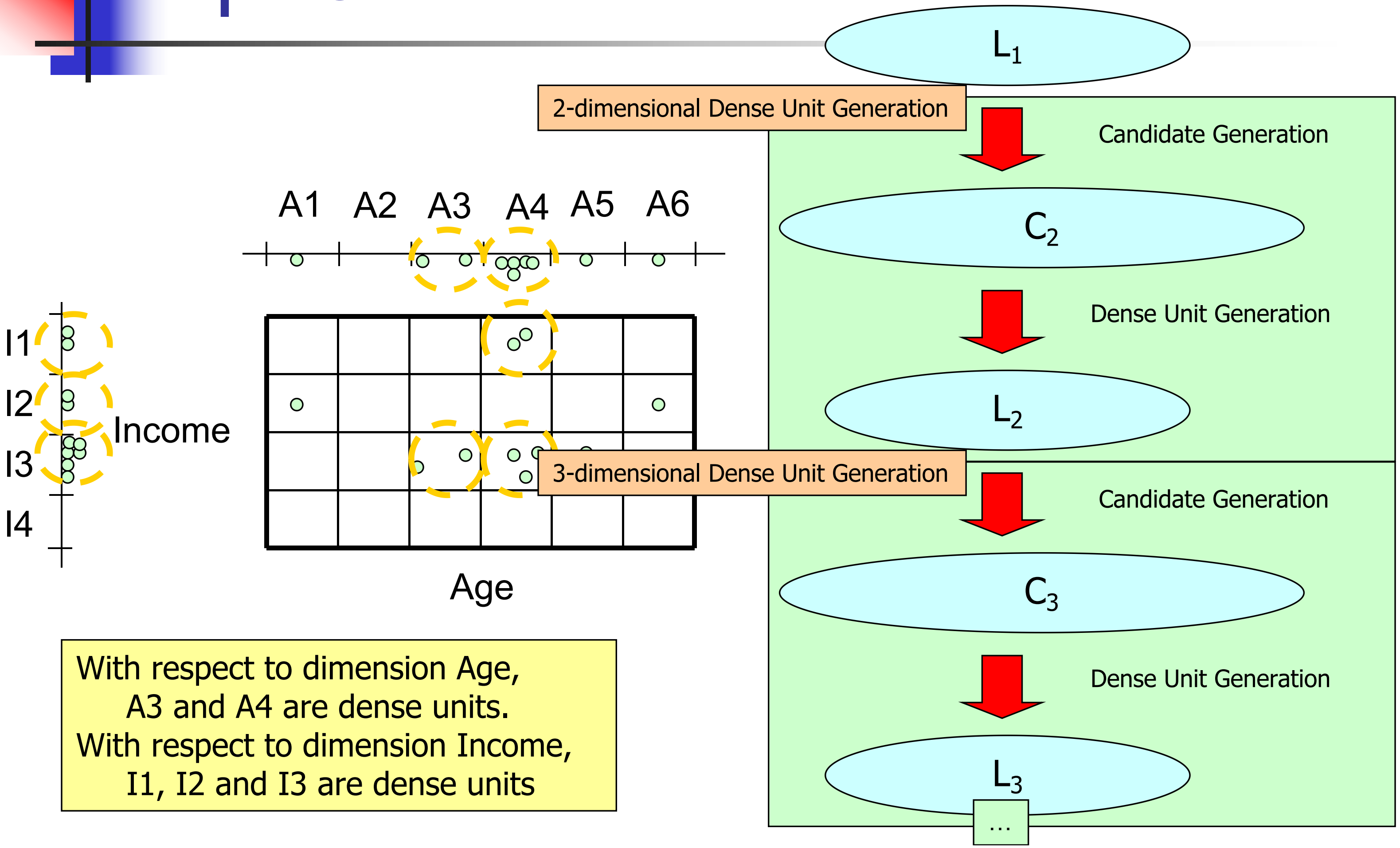
Suppose we want to find all dense units (e.g., dense units with density >= 20%)



With respect to dimension Age,
A3 and A4 are dense units.
With respect to dimension Income,
I1, I2 and I3 are dense units

Apriori

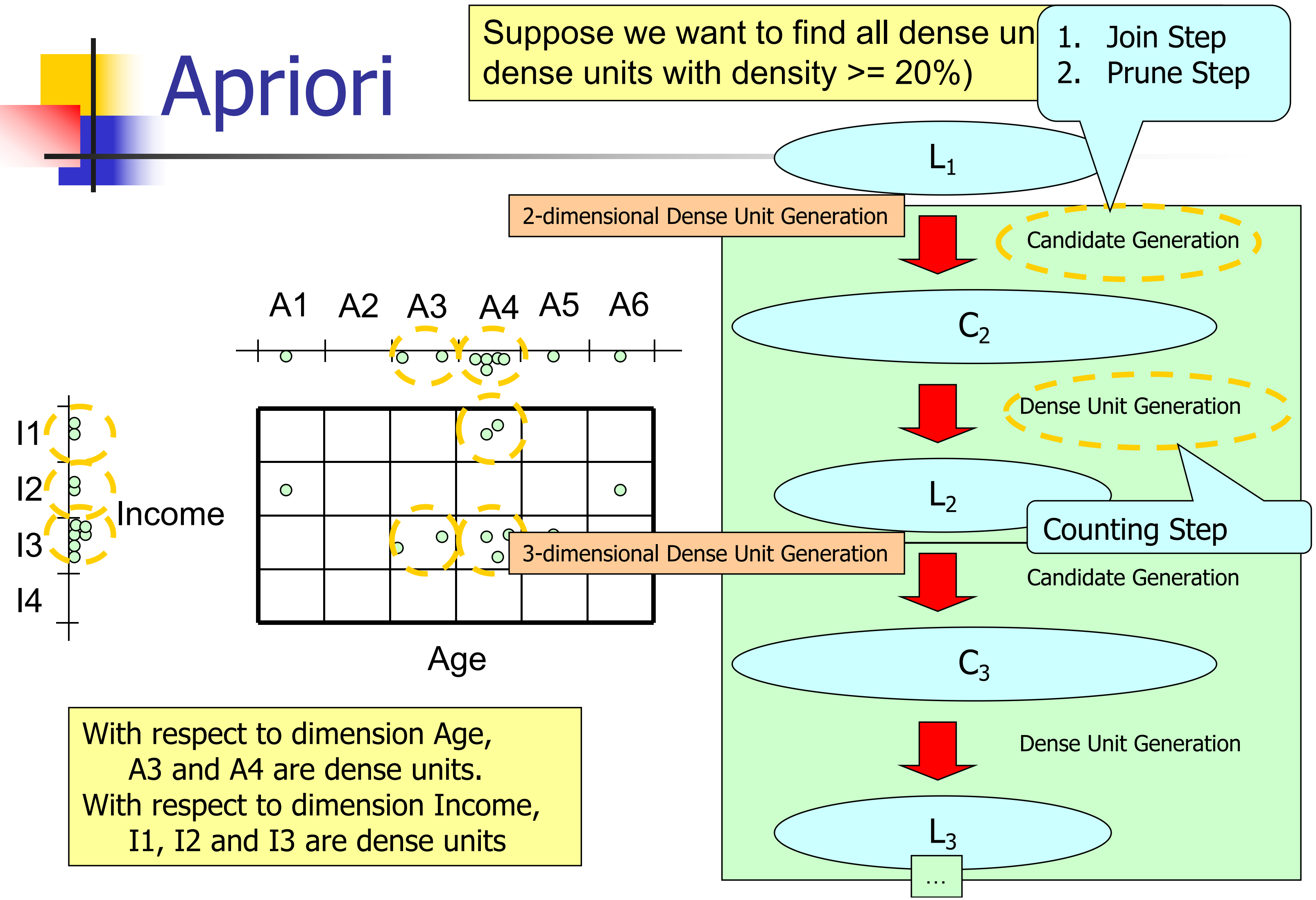
Suppose we want to find all dense units (e.g., dense units with density >= 20%)



With respect to dimension Age, A3 and A4 are dense units.
With respect to dimension Income, I1, I2 and I3 are dense units

Apriori

Suppose we want to find all dense un
dense units with density >= 20%)



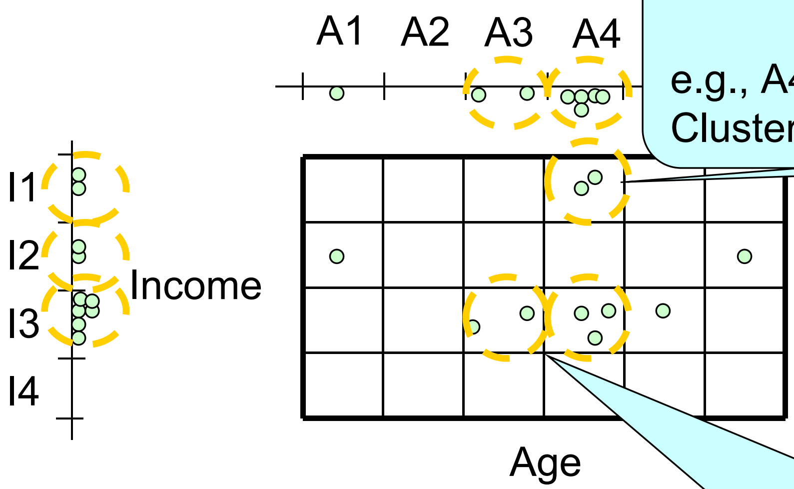
With respect to dimension Age,
A3 and A4 are dense units.
With respect to dimension Income,
I1, I2 and I3 are dense units

Dense Unit-based Method for Subspace Clustering

- **Step 1:** Identify sub-spaces that contain dense units
- **Step 2:** Identify clusters in each sub-spaces that contain dense units

Step 2

Suppose we want to find all dense units (e.g., dense units with density >= 20%)



Cluster 2: A4 and I1

e.g., A4 = 21-25 and I1 = 10k-15k
Cluster 2: Age= 21-25 and Income= 10k-15k

With respect to dimension Age,
A3 and A4 are dense units.
With respect to dimension Incom
I1, I2 and I3 are dense units

Cluster 1: (A3 or A4) and I3

e.g., A3 = 16-20, A4 = 21-25 and I3 = 20k-25k
Cluster 1: Age=16-25 and Income=20k-25k

Subspace Clustering Methods

- Dense Unit-based Method
- Entropy-Based Method
- Transformation-Based Method

Entropy-Based Method for Subspace Clustering

- Entropy
- Problem
  - Good subspace Clustering
- Algorithm
  - Property
  - Apriori

Entropy

- Example
  - Suppose we have a horse race with eight horses taking part.
  - Assume that the probabilities of winning for the eight horses are
  - $\frac{1}{2}$, $\frac{1}{4}$, 1/8, 1/16, 1/64, 1/64, 1/64, 1/64

Entropy

- Suppose we want to send a message to another person indicating which horse won the race. One method is to send a 3 bit string to denote the index of the winning horse

Entropy

- Another method is to use a variable length coding set (i.e. 0, 10, 110, 1110, 111100, 111101, 111110, 111111) to represent the eight horses.
- The average description length is
  - ½ log ½ - ¼ log ¼ - 1/8 log 1/8 – 1/16 log 1/16 – 4 x 1/64 log 1/64
  = 2 bits

Entropy

- The **entropy** is a way to measure the amount of information.
- The smaller the entropy (viewed as the average length of description length in the above example), the more informative we have.

Entropy

- Assume that the probabilities of winning for the eight horses are
- (1/8, 1/8, 1/8, 1/8, 1/8, 1/8, 1/8, 1/8)
- Entropy of the horse race:
- $H(X) = -(1/8 \log 1/8) \times 8$
  $= 3$ bits

Entropy

- Assume that the probabilities of winning for the eight horses are
- (1, 0, 0, 0, 0, 0, 0, 0)
- Entropy of the horse race:
- $H(X) = -1 \log 1 - 7 (0 \log 0)$
  $= 0$ bits

We use the convention that $0 \log 0 = 0$

justified by continuity since $x \log x \rightarrow 0$ as $x \rightarrow 0$

Entropy

- Let A be the set of possible outcomes of random variable X.
- Let p(x) be the probability mass function of the random variable X.
- The entropy H(X) of a discrete random variable X is

$$H(X) = - \sum_{x \in A} p(x) \log p(x)$$

Unit: bit

If the base of log is 2, the unit for entropy is bit.

Entropy

- $H(X) \geq 0$
- Because $0 \leq p(x) \leq 1$

More variables

- When there are more than one variable, we can calculate the **joint entropy** to measure their uncertainty
- X_i: the i-th random variable
- A_i: the set of possible outcomes of X_i
- Entropy:

$$H(X_1, \ldots, X_n) = - \sum_{x_1 \in A_1} \cdots \sum_{x_n \in A_n} p(x_1, \ldots, x_n) \log p(x_1, \ldots, x_n)$$

More variables

$X_1 \backslash X_2$	1	2
1	1/4	½
2	0	1/4

$p(1, 1) = ¼$ $p(1, 2) = ½$ $p(2, 1) = 0$ $p(2, 2) = ¼$

$H(X_1, X_2)$

$= - ¼ \log ¼ - ½ \log ½ - 0 \log 0 – ¼ \log ¼$

$= 1.5$ bits

Entropy-Based Method for Subspace Clustering

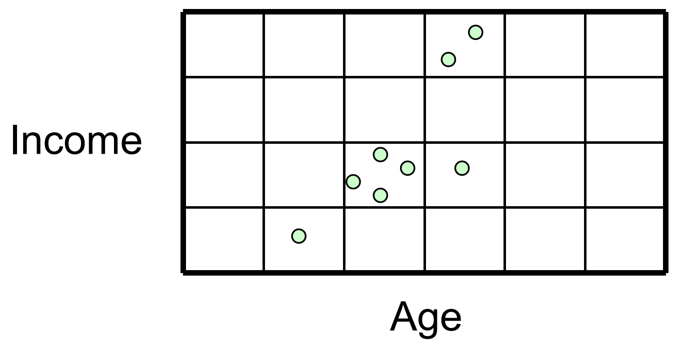
- Entropy
- Problem
  - Good subspace Clustering
- Algorithm
  - Property
  - Apriori

Subspace Clustering

- We divide each dimension into intervals of equal length Δ, so the subspace is partitioned into a grid.

Subspace Clustering

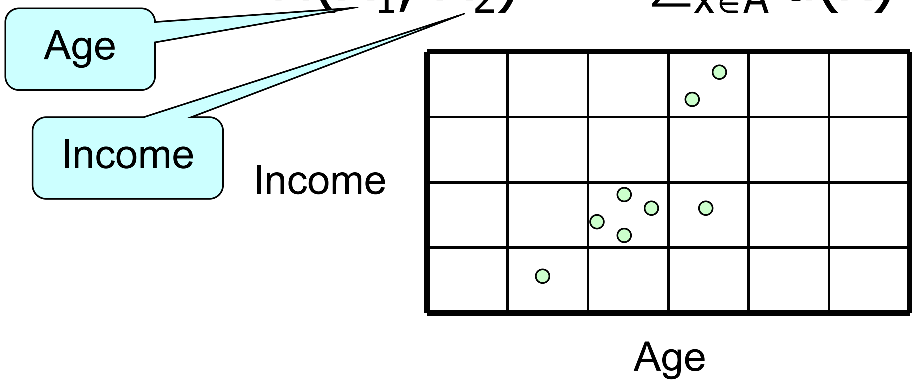
- We divide each dimension into intervals of equal length Δ, so the subspace is partitioned into a grid.



Subspace Clustering

- Let A be the set of all cells.
- d(x) be the density of a cell x in terms of the percentage of data contained in x.
- We can define the entropy to be:

$$H(X_1, X_2) = - \sum_{x \in A} d(x) \log d(x)$$



Subspace Clustering

- Let A be the set of all cells.
- d(x) be the density of a cell x in terms of the percentage of data contained in x.
- We can define the entropy to be:

$$H(X_1, X_2) = - \sum_{x \in A} d(x) \log d(x)$$

Age

Income

Given a parameter ω,
k dimensions (or random variables) are said to have **good clustering** if

$$H(X_1, X_2 \ldots, X_k) <= \omega$$

Subspace Clustering

- **Problem:** We want to find all subspaces with good clustering.

e.g., we want to find sub-spaces with entropy <= 0.2

Given a parameter ω,
k dimensions (or random variables) are said to have **good clustering** if

$$H(X_1, X_2\ldots, X_k) <= \omega$$

Conditional Entropy

- The conditional entropy $H(Y|X)$ is defined as

$$H(Y|X) = \sum_{x \in A} p(x) H(Y|X = x)$$

Conditional Entropy

X\Y	1	2
1	0	¾
2	1/8	1/8

$H(Y|X=1) = 0$ bit

$H(Y|X=2) = 1$ bit

$H(Y|X) = \frac{3}{4} \times H(Y|X=1) + \frac{1}{4} \times H(Y|X = 2)$

$= 0.25$ bit

$H(Y|X) = - \sum_{x \in A} \sum_{y \in B} p(x, y) \log p(y|x)$

Conditional Entropy

- A : a set of possible outcomes of random variable X
- B : a set of possible outcomes of random variable Y
- $H(Y|X) = - \sum_{x \in A} \sum_{y \in B} p(x, y) \log p(y|x)$

$H(Y|X) = - \sum_{x \in A} \sum_{y \in B} p(x, y) \log p(y|x)$

Conditional Entropy

X\Y	1	2
1	0	¾
2	1/8	1/8

$p(Y = 1 | X = 1) = 0$

$p(Y = 2 | X = 1) = 1$

$p(Y = 1 | X = 2) = ½$

$p(Y = 2 | X = 2) = ½$

$$H(Y|X) = - 0 \log 0 - ¾ \log 1 - 1/8 \log ½ - 1/8 \log ½$$
$$= 0.25 \text{ bit}$$

Chain Rule

- $H(X, Y) = H(X) + H(Y \mid X)$

Chain Rule

- $H(X, Y) = H(X) + H(Y \mid X)$

- $H(X_1, \ldots, X_{k-1}, X_k)$
  $= H(X_1, \ldots, X_{k-1}) + H(X_k \mid X_1, \ldots, X_{k-1})$

Entropy-Based Method for Subspace Clustering

- Entropy
- Problem
  - Good subspace Clustering
- Algorithm
  - Property
  - Apriori

Given a parameter ω,
k dimensions (or random variables) are said to have **good clustering** if

$$H(X_1, X_2 \ldots, X_k) \leq \omega$$

Property

- **Lemma:** If a k-dimensional subspace $X_1, \ldots, X_k$ has good clustering, then each of the (k-1)-dimensional projections of this space has also good clustering.

Proof: Since the subspace $X_1, \ldots, X_k$ has good clustering,

$$H(X_1, \ldots, X_k) \leq \omega$$

Consider a (k-1)-dimensional projections, say $X_1, \ldots, X_{k-1}$:

$$\begin{aligned} H(X_1, \ldots, X_{k-1}) &\leq H(X_1, \ldots, X_{k-1}) + H(X_k \mid X_1, \ldots, X_{k-1}) \\ &= H(X_1, \ldots, X_k) \\ &\leq \omega \end{aligned}$$

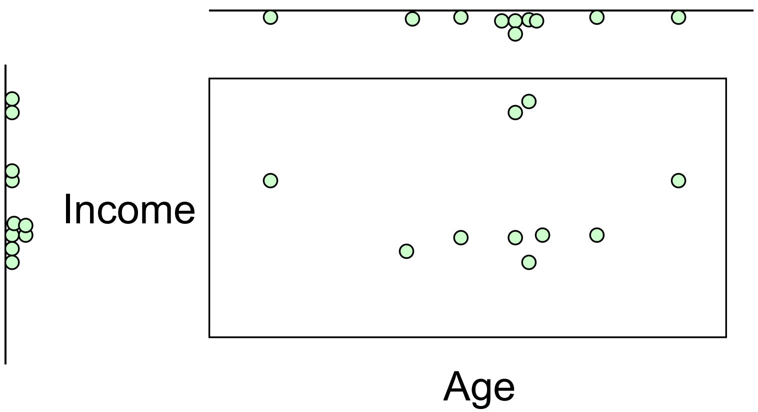
Apriori

Suppose we want to find sub-spaces with entropy <= 0.2

- We can make use of apriori approach to solve the problem

Apriori

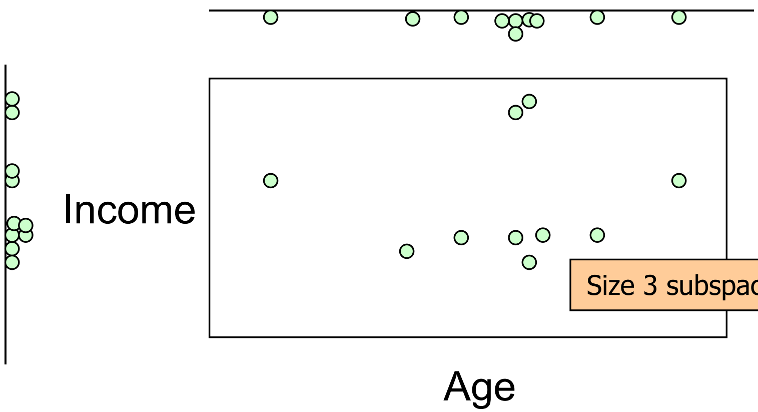
Suppose we want to find sub-spaces with entropy <= 0.2

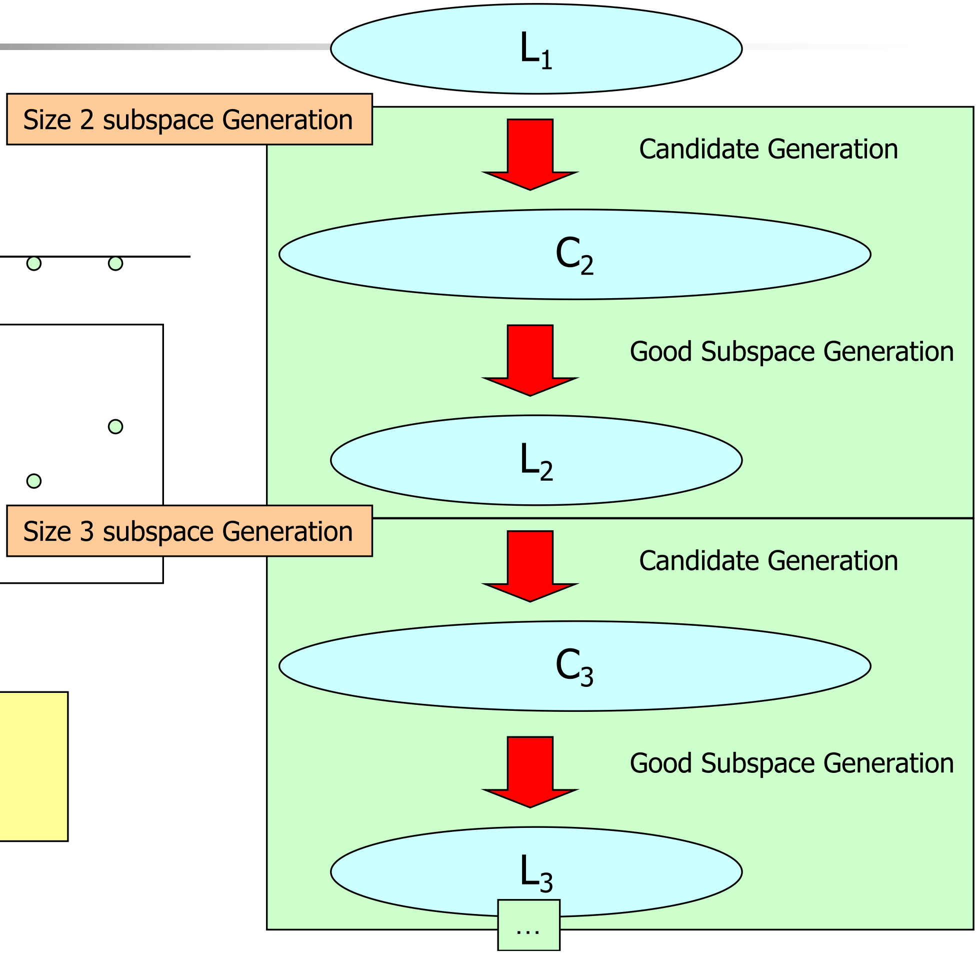


e.g.,
H(Age) = 0.12
H(Income) = 0.08

Apriori

Suppose we want to find sub-spaces with entropy <= 0.2

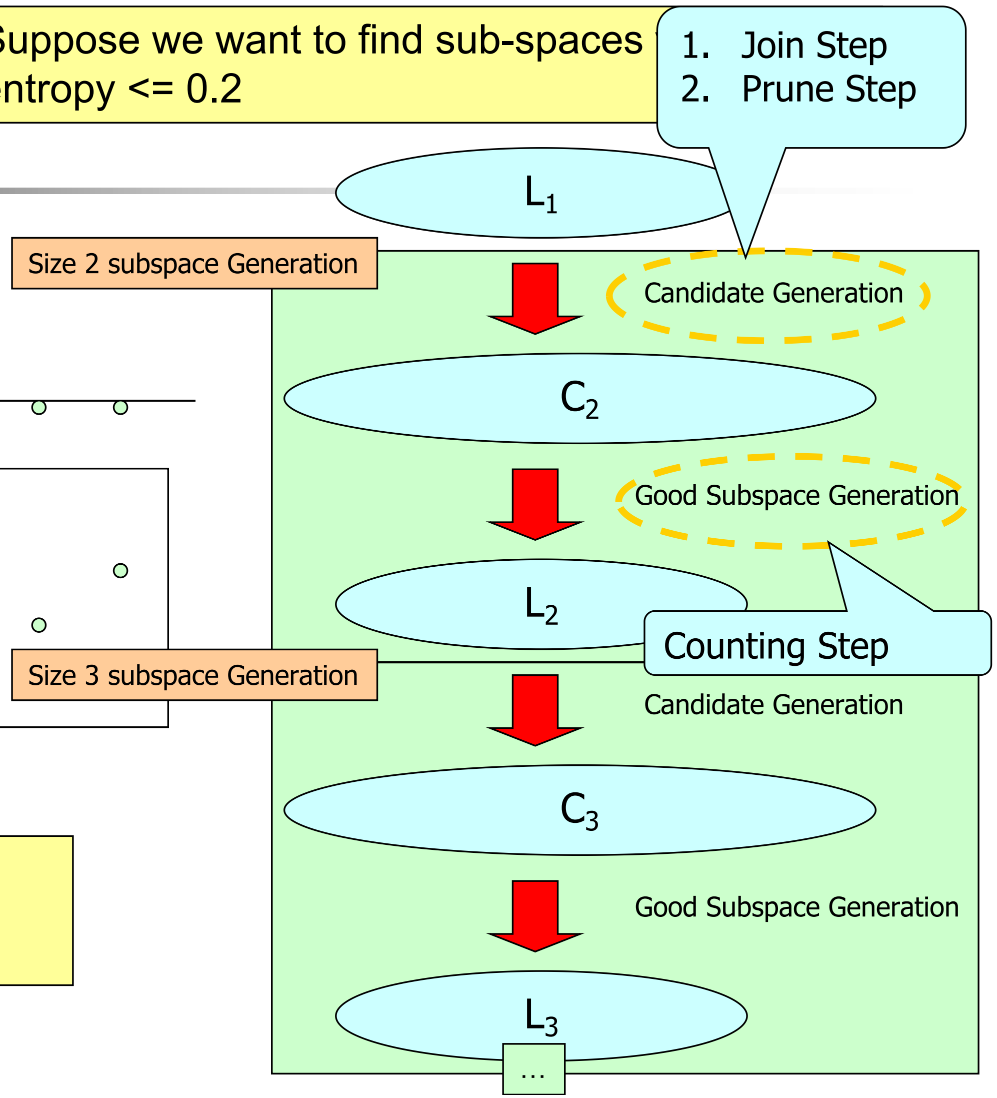


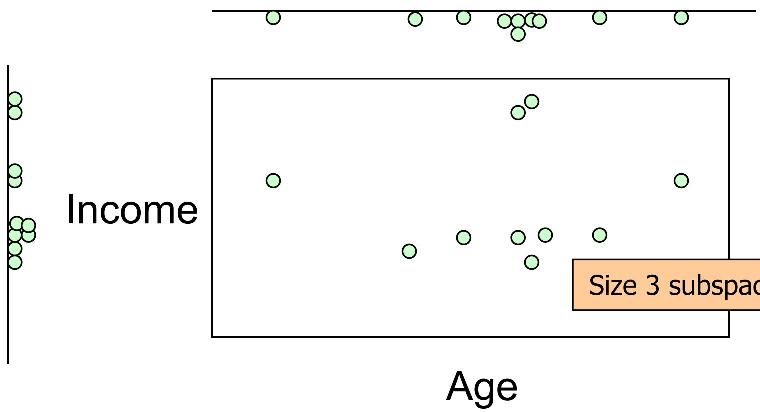


e.g.,
H(Age) = 0.12
H(Income) = 0.08

Apriori

Suppose we want to find sub-spaces
entropy <= 0.2





e.g.,
H(Age) = 0.12
H(Income) = 0.08

Cluster

Suppose we want to find sub-spaces with entropy <= 0.2

- After finding the subspaces with entropy <= 0.2,
- We can find the real clusters by existing methods (e.g., k-mean) in each of the subspaces found.

Subspace Clustering Methods

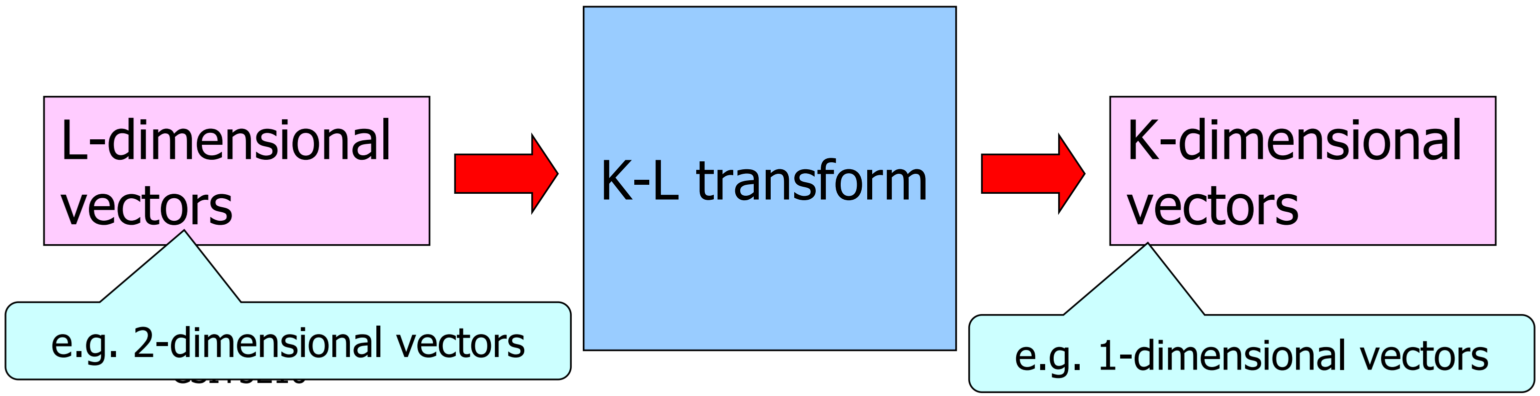
- Dense Unit-based Method
- Entropy-Based Method
- Transformation-Based Method

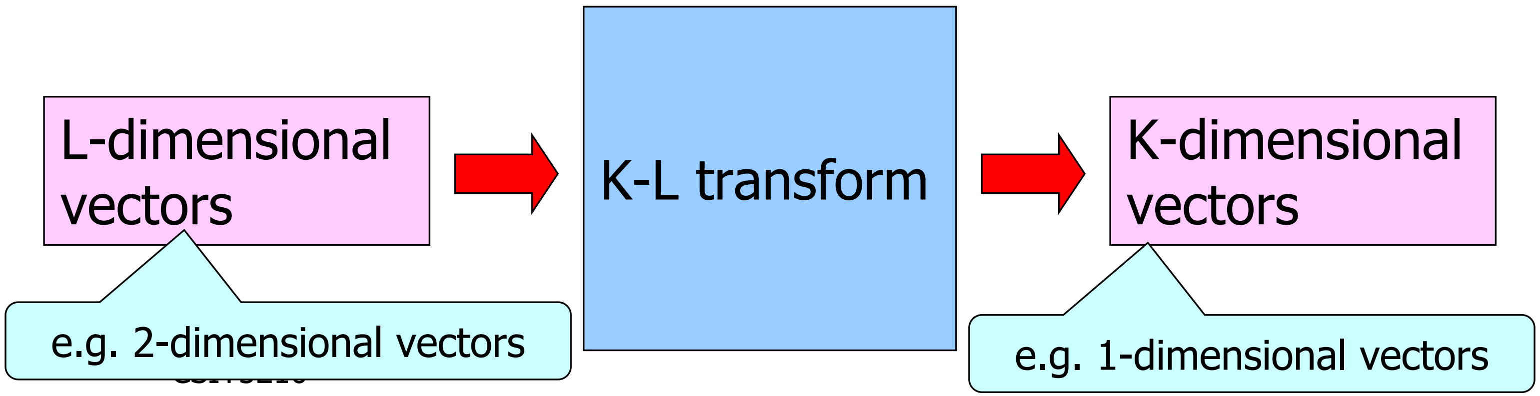
KL-Transform

Karhunen-Loeve Transform

- The two previous approaches find the sub-space in the original dimensions
- KL-Transform "transforms" the data points from the original dimensions into other dimensions

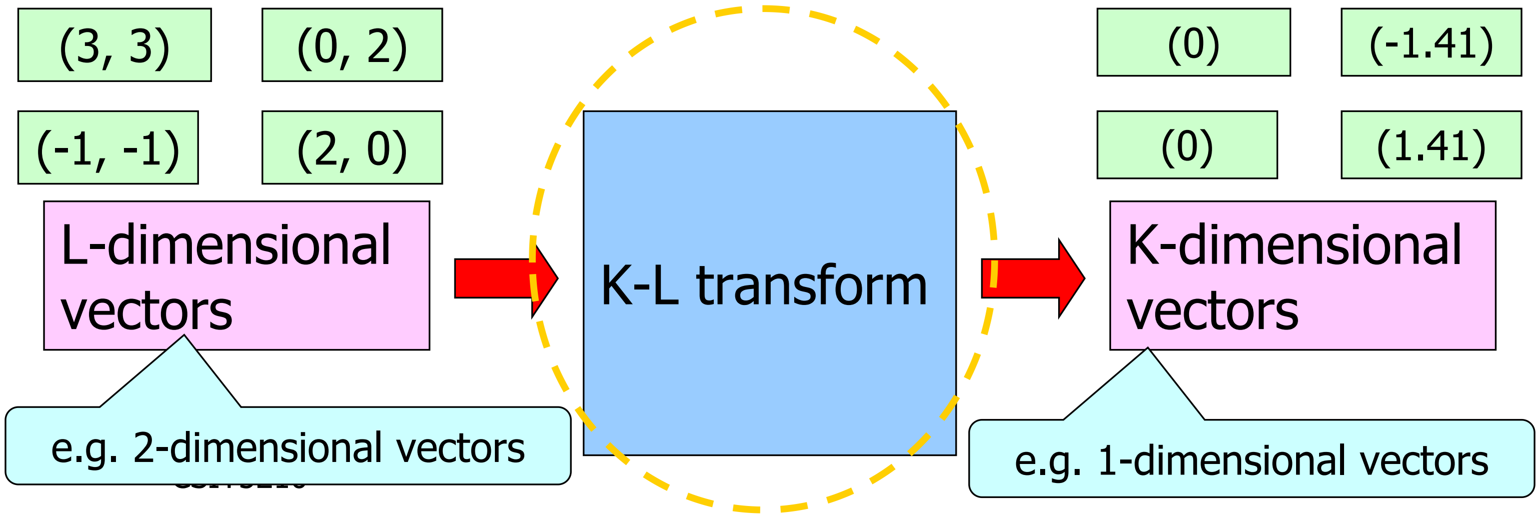
KL-Transform




L-dimensional vectors
K-L transform
K-dimensional vectors
e.g. 2-dimensional vectors
e.g. 1-dimensional vectors

L-dimensional vectors → K-L transform → K-dimensional vectors

Step 1	For each dimension i, calculate the mean e_i (expected value) For each L-dimensional data $\{x_1, x_2, \ldots, x_L\}$ find $\{x_1-e_1, x_2-e_2, \ldots, x_L-e_L\}$
Step 2	Obtain the covariance matrix Σ
Step 3	Find the eigenvalues and eigenvectors of Σ Choose the eigenvectors of unit lengths
Step 4	Arrange the eigenvectors in descending order of the eigenvalues
Step 5	Transform the given L-dimensional vectors by eigenvector matrix
Step 6	For each "transformed" L-dimensional vector, keep only the K values $\{y_1, y_2, \ldots, y_K\}$ corresponding to the smallest K eigenvalues.


(3, 3)
(0, 2)
(-1, -1)
(2, 0)
L-dimensional vectors
e.g. 2-dimensional vectors
K-L transform
(0)
(-1.41)
(0)
(1.41)
K-dimensional vectors
e.g. 1-dimensional vectors

2-dimensional vectors

(3, 3)	(0, 2)
(-1, -1)	(2, 0)

K-L transform

1-dimensional vectors

Step 1

For each dimension i,
calculate the mean e_i (expected value)
For each L-dimensional data $\{x_1, x_2, \ldots, x_L\}$
find $\{x_1-e_1, x_2-e_2, \ldots, x_L-e_L\}$

For **dimension 1**,

mean = (3 + 0 + (-1) + 2)/4 = 4/4 = 1

For **dimension 2**,

mean = (3 + 2 + (-1) + 0)/4 = 4/4 = 1

mean vector = (1, 1)

2-dimensional vectors

(3, 3)	(0, 2)
(-1, -1)	(2, 0)

→ K-L transform → 1-dimensional vectors

Step 1

For each dimension i,
calculate the mean e_i (expected value)
For each L-dimensional data $\{x_1, x_2, \ldots, x_L\}$
find $\{x_1-e_1, x_2-e_2, \ldots, x_L-e_L\}$

For **dimension 1**,

mean =	(3 + 0 + (-1) + 2)/4 =	4/4 = 1

For **dimension 2**,

mean =	(3 + 2 + (-1) + 0)/4 =	4/4 = 1

mean vector = (1, 1)

2-dimensional vectors	
(3, 3)	(0, 2)
(-1, -1)	(2, 0)

→ K-L transform → 1-dimensional vectors

Step 1

For each dimension i,
calculate the mean e_i (expected value)
For each L-dimensional data $\{x_1, x_2, \ldots, x_L\}$
find $\{x_1-e_1, x_2-e_2, \ldots, x_L-e_L\}$

mean vector = (1, 1)

For **data 1**, (3, 3)
Difference from mean vector = (3-1, 3-1) = (2, 2)

For **data 2**, (0, 2)
Difference from mean vector = (0-1, 2-1) = (-1, 1)

For **data 3**, (-1, -1)
Difference from mean vector = (-1-1, -1-1) = (-2, -2)

For **data 4**, (2, 0)
Difference from mean vector = (2-1, 0-1) = (1, -1)

2-dimensional vectors → K-L transform → 1-dimensional vectors

(3, 3)	(0, 2)
(-1, -1)	(2, 0)

Step 1

mean vector = (1, 1)

For **data 1**, (3, 3)

Difference from mean vector = (2, 2)

For **data 2**, (0, 2)

Difference from mean vector = (-1, 1)

For **data 3**, (-1, -1)

Difference from mean vector = (-2, -2)

For **data 4**, (2, 0)

Difference from mean vector = (1, -1)

2-dimensional vectors

(3, 3)	(0, 2)
(-1, -1)	(2, 0)

→ K-L transform → 1-dimensional vectors

Step 1 mean vector = (1, 1)

For **data 1**,	Difference from mean vector = (2, 2)
For **data 2**,	Difference from mean vector = (-1, 1)
For **data 3**,	Difference from mean vector = (-2, -2)
For **data 4**,	Difference from mean vector = (1, -1)

Step 2 Obtain the covariance matrix Σ

$$Y=\begin{pmatrix} 2 & -1 & -2 & 1 \\ 2 & 1 & -2 & -1 \end{pmatrix}$$

$$\Sigma = \frac{1}{4}YY^T= \frac{1}{4}\begin{pmatrix} 2 & -1 & -2 & 1 \\ 2 & 1 & -2 & -1 \end{pmatrix}\begin{pmatrix} 2 & 2 \\ -1 & 1 \\ -2 & -2 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 5/2 & 3/2 \\ 3/2 & 5/2 \end{pmatrix}$$

2-dimensional vectors → K-L transform → 1-dimensional vectors

(3, 3)	(0, 2)
(-1, -1)	(2, 0)

Step 1: mean vector = (1, 1)

For **data 1**,	Difference from mean vector = (2, 2)
For **data 2**,	Difference from mean vector = (-1, 1)
For **data 3**,	Difference from mean vector = (-2, -2)
For **data 4**,	Difference from mean vector = (1, -1)

Step 2

$$Y=\begin{pmatrix} 2 & -1 & -2 & 1 \\ 2 & 1 & -2 & -1 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} 5/2 & 3/2 \\ 3/2 & 5/2 \end{pmatrix}$$

2-dimensional vectors

(3, 3)	(0, 2)
(-1, -1)	(2, 0)

→ K-L transform → 1-dimensional vectors

Step 1 mean vector = (1, 1)

For **data 1**,	Difference from mean vector = (2, 2)
For **data 2**,	Difference from mean vector = (-1, 1)
For **data 3**,	Difference from mean vector = (-2, -2)
For **data 4**,	Difference from mean vector = (1, -1)

Step 2 $\Sigma = \begin{pmatrix} 5/2 & 3/2 \\ 3/2 & 5/2 \end{pmatrix}$

2-dimensional vectors

(3, 3)	(0, 2)
(-1, -1)	(2, 0)

K-L transform

1-dimensional vectors

Step 2

$$\Sigma = \begin{pmatrix} 5/2 & 3/2 \\ 3/2 & 5/2 \end{pmatrix}$$

2-dimensional vectors

(3, 3)	(0, 2)
(-1, -1)	(2, 0)

K-L transform

1-dimensional vectors

Step 2

$$\Sigma = \begin{pmatrix} 5/2 & 3/2 \\ 3/2 & 5/2 \end{pmatrix}$$

Step 3

Find the eigenvalues and eigenvectors of Σ
Choose the eigenvectors of unit lengths

$$\begin{vmatrix} 5/2-\lambda & 3/2 \\ 3/2 & 5/2-\lambda \end{vmatrix} = 0$$

$$(5/2 - \lambda)^2 - (3/2)^2 = 0$$

$$25/4 - 5\lambda + \lambda^2 - 9/4 = 0$$

$$4 - 5\lambda + \lambda^2 = 0$$

$$(\lambda - 4)(\lambda - 1) = 0$$

$$\lambda = 4 \quad \text{or} \quad \lambda = 1$$

2-dimensional vectors	
(3, 3)	(0, 2)
(-1, -1)	(2, 0)

→ K-L transform → 1-dimensional vectors

Step 2

$$\Sigma = \begin{pmatrix} 5/2 & 3/2 \\ 3/2 & 5/2 \end{pmatrix}$$

Step 3

$$\begin{vmatrix} 5/2-\lambda & 3/2 \\ 3/2 & 5/2-\lambda \end{vmatrix} = 0$$

$$(5/2 - \lambda)^2 - (3/2)^2 = 0$$

$$25/4 - 5\lambda + \lambda^2 - 9/4 = 0$$

$$4 - 5\lambda + \lambda^2 = 0$$

$$(\lambda - 4)(\lambda - 1) = 0$$

$$\lambda = 4 \quad \text{or} \quad \lambda = 1$$

2-dimensional vectors

(3, 3)	(0, 2)
(-1, -1)	(2, 0)

→ K-L transform → 1-dimensional vectors

Step 2 $\Sigma = \begin{pmatrix} 5/2 & 3/2 \\ 3/2 & 5/2 \end{pmatrix}$

Step 3 $\lambda = 4$ or $\lambda = 1$

When $\lambda = 4$

$$\begin{pmatrix} 5/2-4 & 3/2 \\ 3/2 & 5/2-4 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} -3/2 & 3/2 \\ 3/2 & -3/2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$x_1 - x_2 = 0$

$x_1 = x_2$

$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a \\ a \end{pmatrix}$ where $a \in R$

We choose the eigenvector of unit length

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \sqrt{1/2} \\ \sqrt{1/2} \end{pmatrix}$$

2-dimensional vectors

(3, 3)	(0, 2)
(-1, -1)	(2, 0)

K-L transform

1-dimensional vectors

Step 2

$$\sum = \begin{pmatrix} 5/2 & 3/2 \\ 3/2 & 5/2 \end{pmatrix}$$

Step 3

$\lambda = 4$ or $\lambda = 1$

When $\lambda = 4$

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \sqrt{1/2} \\ \sqrt{1/2} \end{pmatrix}$$

2-dimensional vectors → K-L transform → 1-dimensional vectors

(3, 3)	(0, 2)
(-1, -1)	(2, 0)

Step 2

$$\Sigma = \begin{pmatrix} 5/2 & 3/2 \\ 3/2 & 5/2 \end{pmatrix}$$

Step 3

$\lambda = 4$ or $\lambda = 1$

When $\lambda = 4$

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \sqrt{1/2} \\ \sqrt{1/2} \end{pmatrix}$$

When $\lambda = 1$

$$\begin{pmatrix} 5/2-1 & 3/2 \\ 3/2 & 5/2-1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 3/2 & 3/2 \\ 3/2 & 3/2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$x_1 + x_2 = 0$

$x_1 = -x_2$

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a \\ -a \end{pmatrix} \quad \text{where} \quad a \in R$$

We choose the eigenvector of unit length

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \sqrt{1/2} \\ -\sqrt{1/2} \end{pmatrix}$$

2-dimensional vectors

(3, 3)	(0, 2)
(-1, -1)	(2, 0)

K-L transform

1-dimensional vectors

Step 2 $\Sigma = \begin{pmatrix} 5/2 & 3/2 \\ 3/2 & 5/2 \end{pmatrix}$

Step 3 $\lambda = 4$ or $\lambda = 1$

When $\lambda = 4$ $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \sqrt{1/2} \\ \sqrt{1/2} \end{pmatrix}$

When $\lambda = 1$

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \sqrt{1/2} \\ -\sqrt{1/2} \end{pmatrix}$$

2-dimensional vectors

(3, 3)	(0, 2)
(-1, -1)	(2, 0)

K-L transform

1-dimensional vectors

Step 2 $\sum = \begin{pmatrix} 5/2 & 3/2 \\ 3/2 & 5/2 \end{pmatrix}$

Step 3 $\lambda = 4$ or $\lambda = 1$

When $\lambda = 4$ $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \sqrt{1/2} \\ \sqrt{1/2} \end{pmatrix}$

When $\lambda = 1$ $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \sqrt{1/2} \\ -\sqrt{1/2} \end{pmatrix}$

2-dimensional vectors → K-L transform → 1-dimensional vectors

(3, 3)	(0, 2)
(-1, -1)	(2, 0)

Step 3 $\lambda = 4$ or $\lambda = 1$

When $\lambda = 4$ $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \sqrt{1/2} \\ \sqrt{1/2} \end{pmatrix}$

When $\lambda = 1$ $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \sqrt{1/2} \\ -\sqrt{1/2} \end{pmatrix}$

Step 4 Arrange the eigenvectors in descending order of the eigenvalues

$$\Phi = \begin{pmatrix} \sqrt{1/2} & \sqrt{1/2} \\ \sqrt{1/2} & -\sqrt{1/2} \end{pmatrix}$$

2-dimensional vectors → K-L transform → 1-dimensional vectors

(3, 3)	(0, 2)
(-1, -1)	(2, 0)

Step 4	Arrange the eigenvectors in descending order of the eigenvalues

$$\Phi = \begin{pmatrix} \sqrt{1/2} & \sqrt{1/2} \\ \sqrt{1/2} & -\sqrt{1/2} \end{pmatrix}$$

2-dimensional vectors → K-L transform → 1-dimensional vectors

(3, 3)	(0, 2)
(-1, -1)	(2, 0)

Step 4

$$\Phi = \begin{pmatrix} \sqrt{1/2} & \sqrt{1/2} \\ \sqrt{1/2} & -\sqrt{1/2} \end{pmatrix}$$

Step 5 Transform the given L-dimensional vectors by eigenvector matrix

$$Y = \Phi^T X$$

$$Y = \begin{pmatrix} \sqrt{1/2} & \sqrt{1/2} \\ \sqrt{1/2} & -\sqrt{1/2} \end{pmatrix} X$$

2-dimensional vectors	
(3, 3)	(0, 2)
(-1, -1)	(2, 0)

→ K-L transform → 1-dimensional vectors

Step 4

$$\Phi = \begin{pmatrix} \sqrt{1/2} & \sqrt{1/2} \\ \sqrt{1/2} & -\sqrt{1/2} \end{pmatrix}$$

Step 5 Transform the given L-dimensional vectors by eigenvector matrix

$$Y = \begin{pmatrix} \sqrt{1/2} & \sqrt{1/2} \\ \sqrt{1/2} & -\sqrt{1/2} \end{pmatrix} X$$

2-dimensional vectors

(3, 3)	(0, 2)
(-1, -1)	(2, 0)

K-L transform

1-dimensional vectors

Step 4

$$\Phi = \begin{pmatrix} \sqrt{1/2} & \sqrt{1/2} \\ \sqrt{1/2} & -\sqrt{1/2} \end{pmatrix}$$

$$Y = \begin{pmatrix} \sqrt{1/2} & \sqrt{1/2} \\ \sqrt{1/2} & -\sqrt{1/2} \end{pmatrix} X$$

Step 5 Transform the given L-dimensional vectors by eigenvector matrix

2-dimensional vectors	
(3, 3)	(0, 2)
(-1, -1)	(2, 0)

→ **K-L transform** → **1-dimensional vectors**

Step 4

$$\Phi = \begin{pmatrix} \sqrt{1/2} & \sqrt{1/2} \\ \sqrt{1/2} & -\sqrt{1/2} \end{pmatrix}$$

$$Y = \begin{pmatrix} \sqrt{1/2} & \sqrt{1/2} \\ \sqrt{1/2} & -\sqrt{1/2} \end{pmatrix} X$$

Step 5 Transform the given L-dimensional vectors by eigenvector matrix

For **data 1**, (3, 3)

$$Y = \begin{pmatrix} \sqrt{1/2} & \sqrt{1/2} \\ \sqrt{1/2} & -\sqrt{1/2} \end{pmatrix} \begin{pmatrix} 3 \\ 3 \end{pmatrix} = \begin{pmatrix} 4.24 \\ 0 \end{pmatrix}$$

For **data 2**, (0, 2)

$$Y = \begin{pmatrix} \sqrt{1/2} & \sqrt{1/2} \\ \sqrt{1/2} & -\sqrt{1/2} \end{pmatrix} \begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 1.41 \\ -1.41 \end{pmatrix}$$

For **data 3**, (-1, -1)

$$Y = \begin{pmatrix} \sqrt{1/2} & \sqrt{1/2} \\ \sqrt{1/2} & -\sqrt{1/2} \end{pmatrix} \begin{pmatrix} -1 \\ -1 \end{pmatrix} = \begin{pmatrix} -1.41 \\ 0 \end{pmatrix}$$

For **data 4**, (2, 0)

$$Y = \begin{pmatrix} \sqrt{1/2} & \sqrt{1/2} \\ \sqrt{1/2} & -\sqrt{1/2} \end{pmatrix} \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 1.41 \\ 1.41 \end{pmatrix}$$

2-dimensional vectors

(3, 3)	(0, 2)
(-1, -1)	(2, 0)

→ K-L transform → 1-dimensional vectors

Step 4

$$\Phi = \begin{pmatrix} \sqrt{1/2} & \sqrt{1/2} \\ \sqrt{1/2} & -\sqrt{1/2} \end{pmatrix}$$

$$Y = \begin{pmatrix} \sqrt{1/2} & \sqrt{1/2} \\ \sqrt{1/2} & -\sqrt{1/2} \end{pmatrix} X$$

Step 5

Transform the given L-dimensional vectors by eigenvector matrix

For **data 1**, (3, 3)

$$= \begin{pmatrix} 4.24 \\ 0 \end{pmatrix}$$

For **data 3**, (-1, -1)

$$Y = \begin{pmatrix} \sqrt{1/2} & \sqrt{1/2} \\ \sqrt{1/2} & -\sqrt{1/2} \end{pmatrix} \begin{pmatrix} -1 \\ -1 \end{pmatrix} = \begin{pmatrix} -1.41 \\ 0 \end{pmatrix}$$

For **data 2**, (0, 2)

$$Y = \begin{pmatrix} \sqrt{1/2} & \sqrt{1/2} \\ \sqrt{1/2} & -\sqrt{1/2} \end{pmatrix} \begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 1.41 \\ -1.41 \end{pmatrix}$$

For **data 4**, (2, 0)

$$Y = \begin{pmatrix} \sqrt{1/2} & \sqrt{1/2} \\ \sqrt{1/2} & -\sqrt{1/2} \end{pmatrix} \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 1.41 \\ 1.41 \end{pmatrix}$$

2-dimensional vectors	
(3, 3)	(0, 2)
(-1, -1)	(2, 0)

→ K-L transform → 1-dimensional vectors

Step 4	$\Phi = \begin{pmatrix} \sqrt{1/2} & \sqrt{1/2} \\ \sqrt{1/2} & -\sqrt{1/2} \end{pmatrix}$

$$Y = \begin{pmatrix} \sqrt{1/2} & \sqrt{1/2} \\ \sqrt{1/2} & -\sqrt{1/2} \end{pmatrix} X$$

Step 5 Transform the given L-dimensional vectors by eigenvector matrix

For **data 1**,	(3, 3)	→	(4.24, 0)
For **data 2**,	(0, 2)	→	(1.41, -1.41)
For **data 3**,	(-1, -1)	→	(-1.41, 0)
For **data 4**,	(2, 0)	→	(1.41, 1.41)

2-dimensional vectors

(3, 3)	(0, 2)
(-1, -1)	(2, 0)

K-L transform

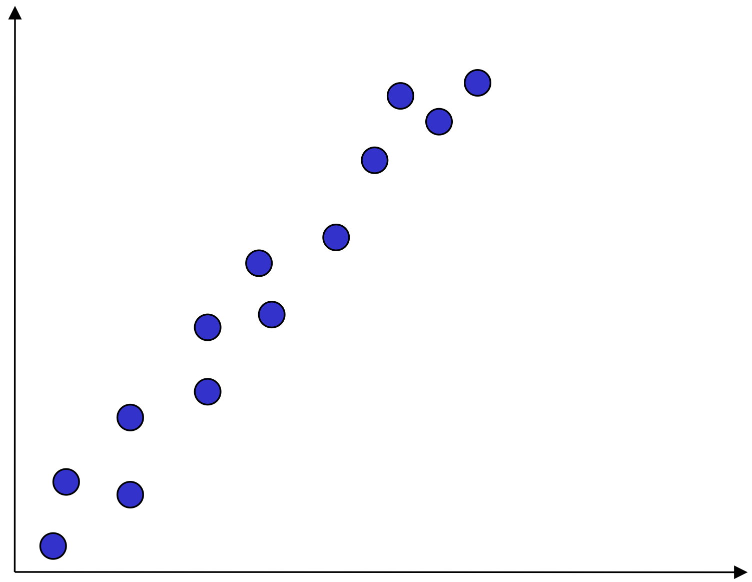
1-dimensional vectors

Step 6 For each “transformed” L-dimensional vector, keep only the K values $\{y_1, y_2, \ldots, y_K\}$ corresponding to the smallest k eigenvalues.

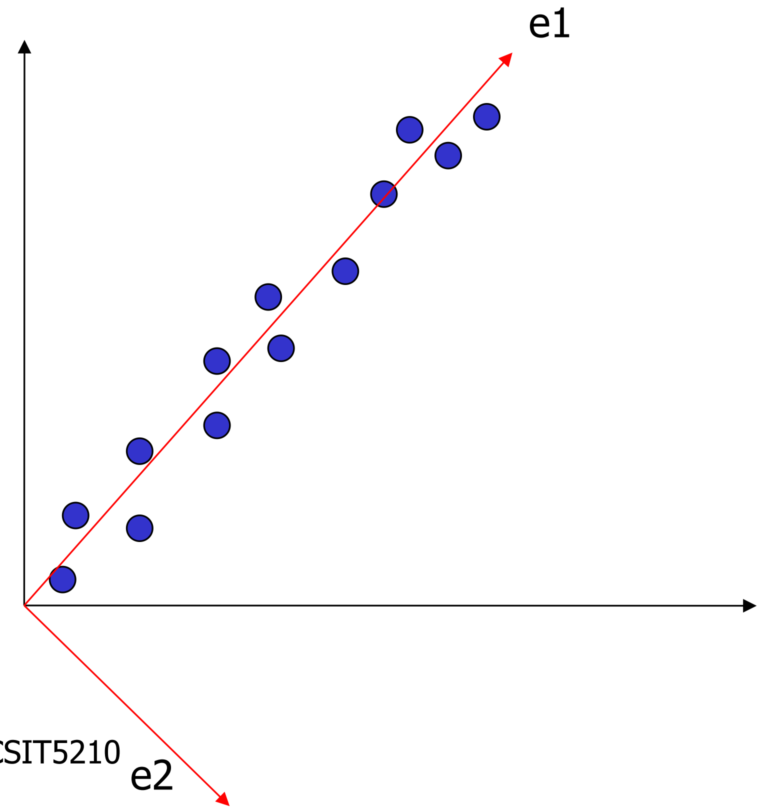
For **data 1**,	(3, 3)	(4.24, 0)	(0)
For **data 2**,	(0, 2)	(1.41, -1.41)	(-1.41)
For **data 3**,	(-1, -1)	(-1.41, 0)	(0)
For **data 4**,	(2, 0)	(1.41, 1.41)	(1.41)

- Why do we need to do this KL-transform?
- Why do we choose the eigenvectors with the smallest eigenvalues?

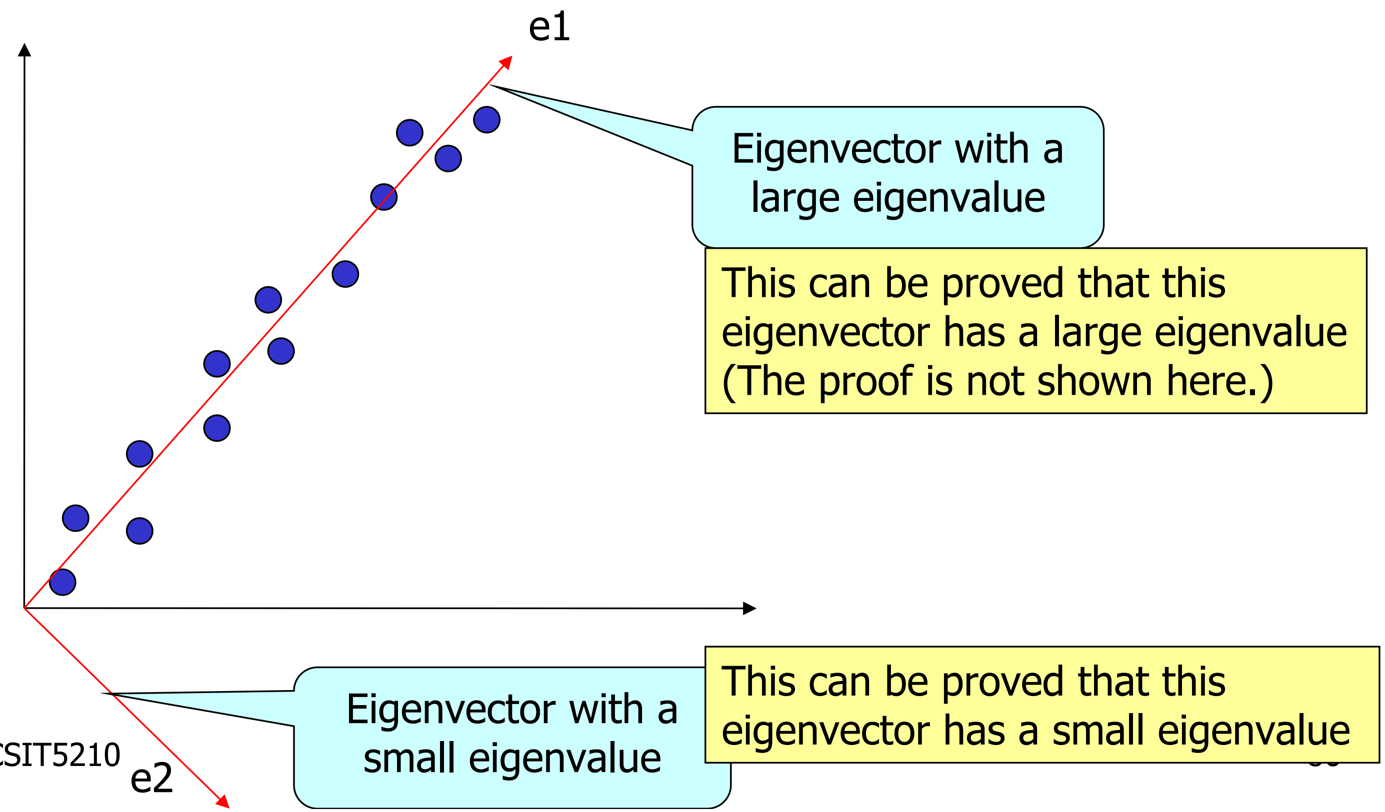
Suppose we have the following data set



According to the data, we find the following eigenvectors (marked in red)



According to the data, we find the following eigenvectors (marked in red)



Consider two cases

- Case 1
  - Consider that the data points are projected on e1

It corresponds to dimension reduction.

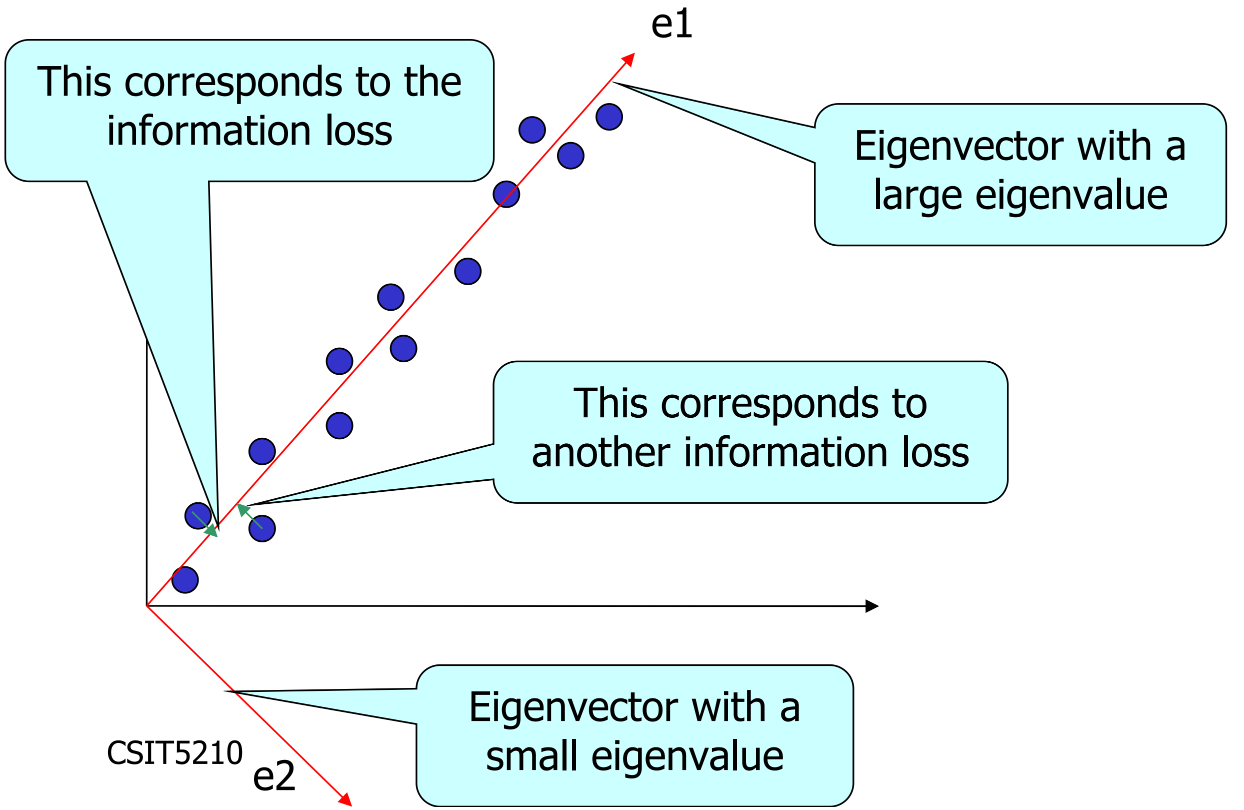
- Case 2
  - Consider that the data points are projected on e2

It corresponds to subspace clustering.

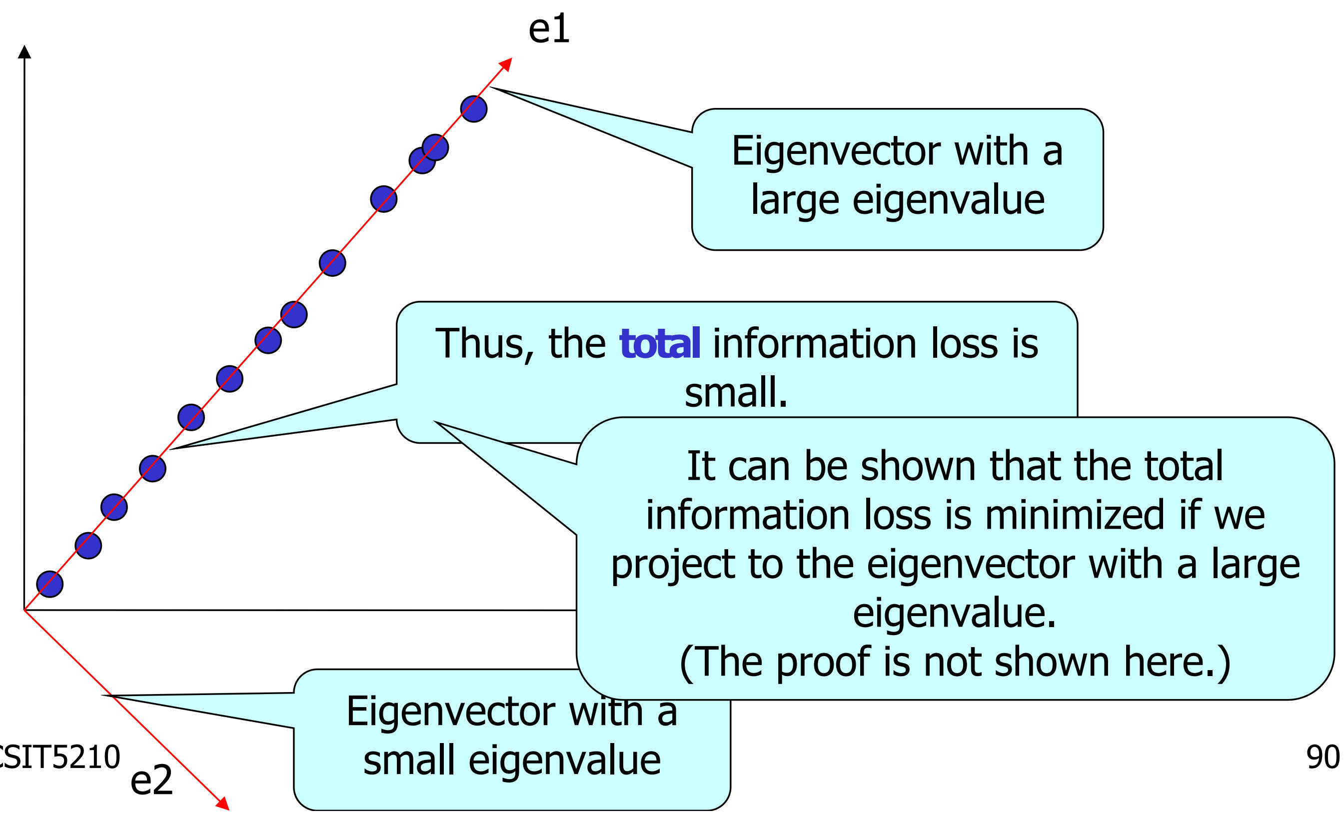
Case 1

- Consider that the data points are projected on e1

Suppose all data points are projected on vector e1



After all data points are projected on vector e1



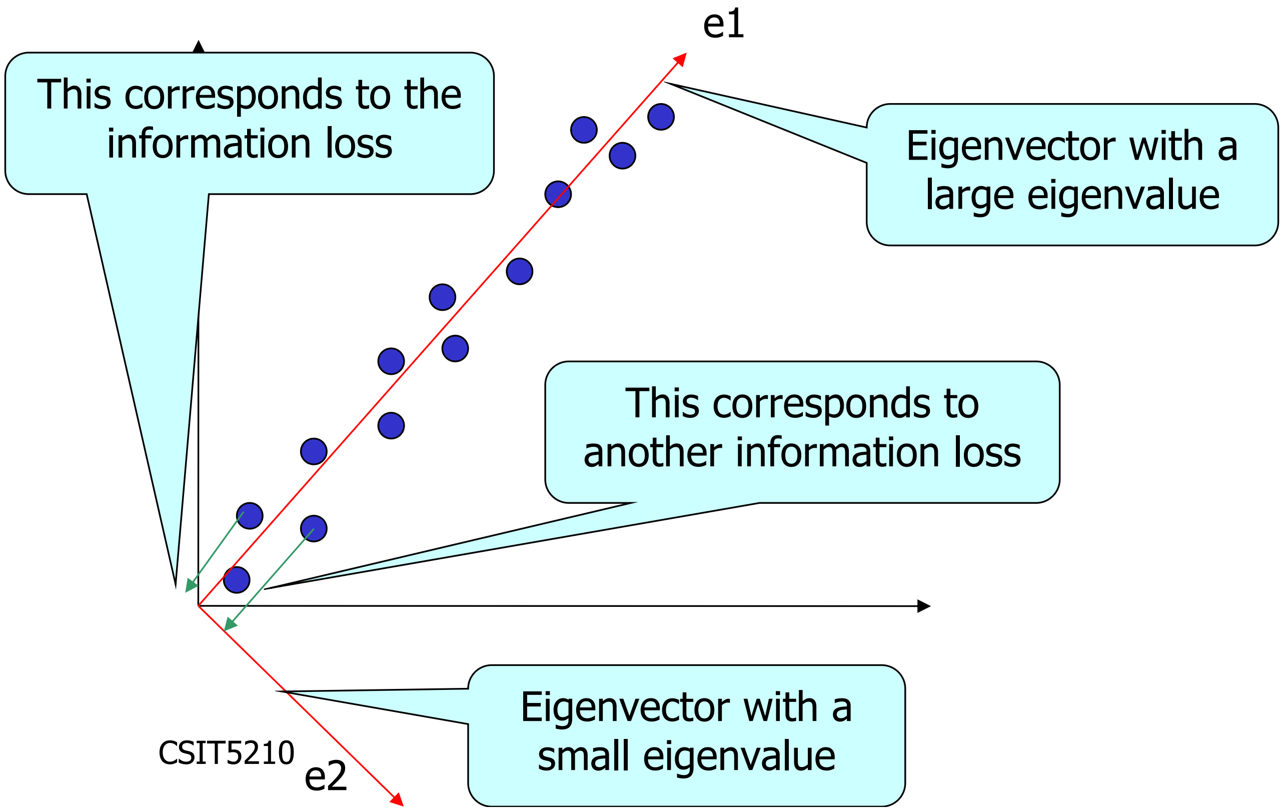
Objective of Dimension Reduction

- The objective of Dimension Reduction
  - To **reduce** the total number of dimensions
  - At the same time, we want to keep information as much as possible.
    (i.e., **minimize** the information loss)
- In our example,
  - We reduce from two dimensions to one dimension
  - The eigenvector with a large eigenvector corresponds to this dimension
  - After we adopt this dimension, we can minimize the information loss

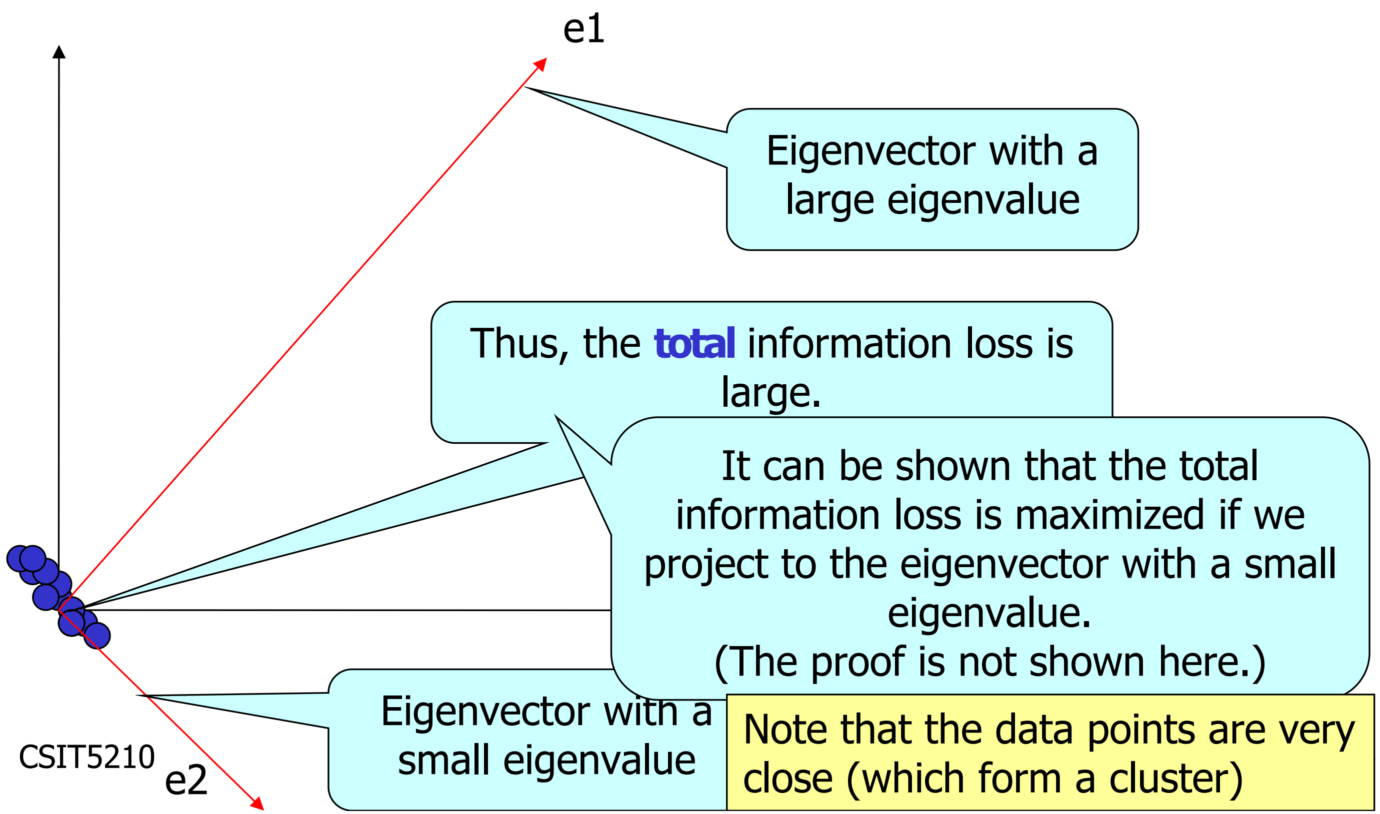
Case 2

- Consider that the data points are projected on e2

Suppose all data points are projected on vector e2



After all data points are projected on vector e2



Objective of Subspace Clustering (KL-Transform)

- The objective of Subspace clustering
  - To **reduce** the total number of dimensions
  - At the same time, we want to find a cluster
    - A cluster is a group of "close" data points
    - This means that, after the data points are transformed, the data points are very close.
    - In KL-transform, you can see that the information loss is **maximized**.
- In our example,
  - We reduce from two dimensions to one dimension
  - The eigenvector with a small eigenvector corresponds to this dimension
  - After we adopt this dimension, we can maximize the information loss