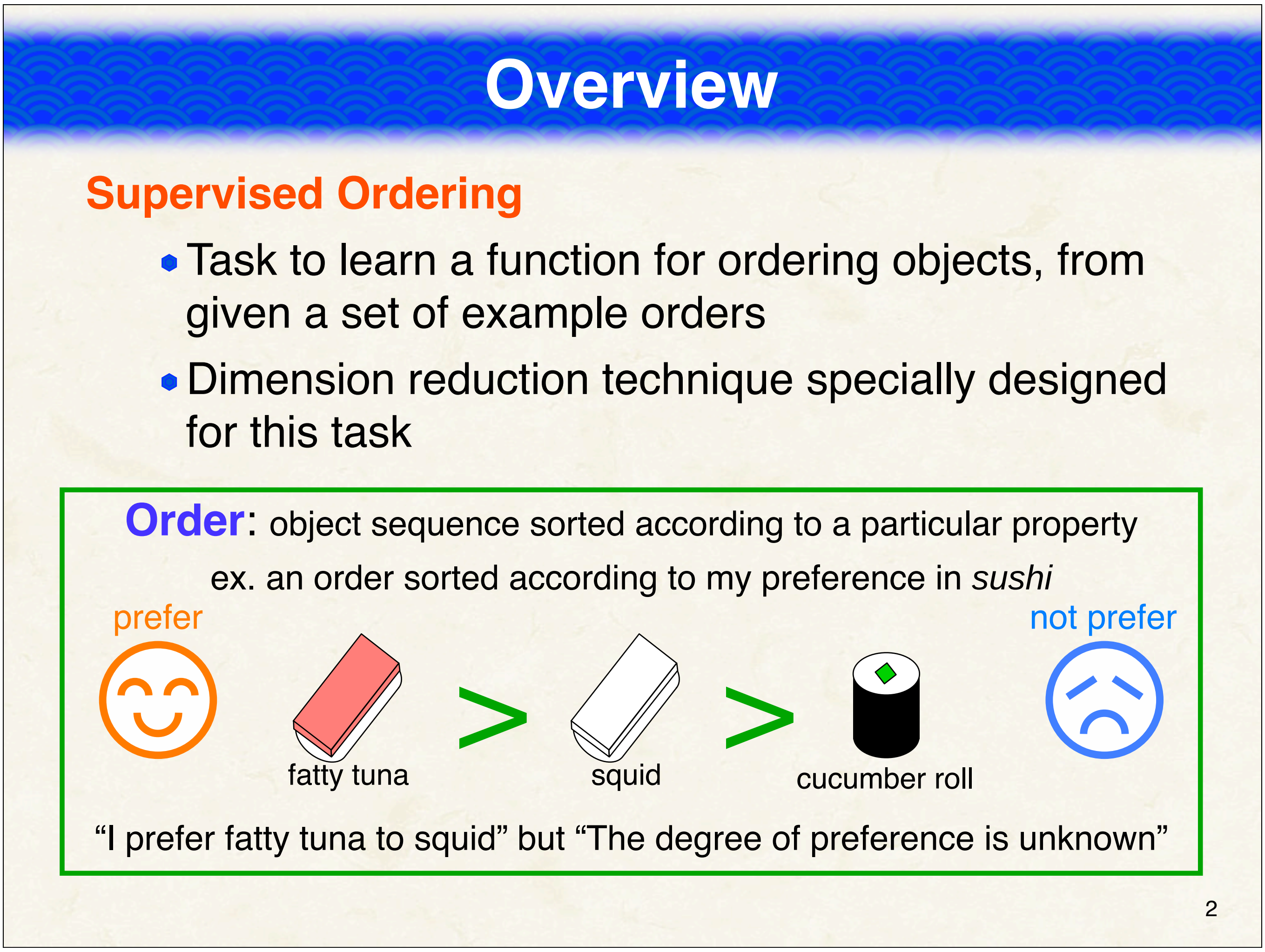
Dimension Reduction for Supervised Ordering

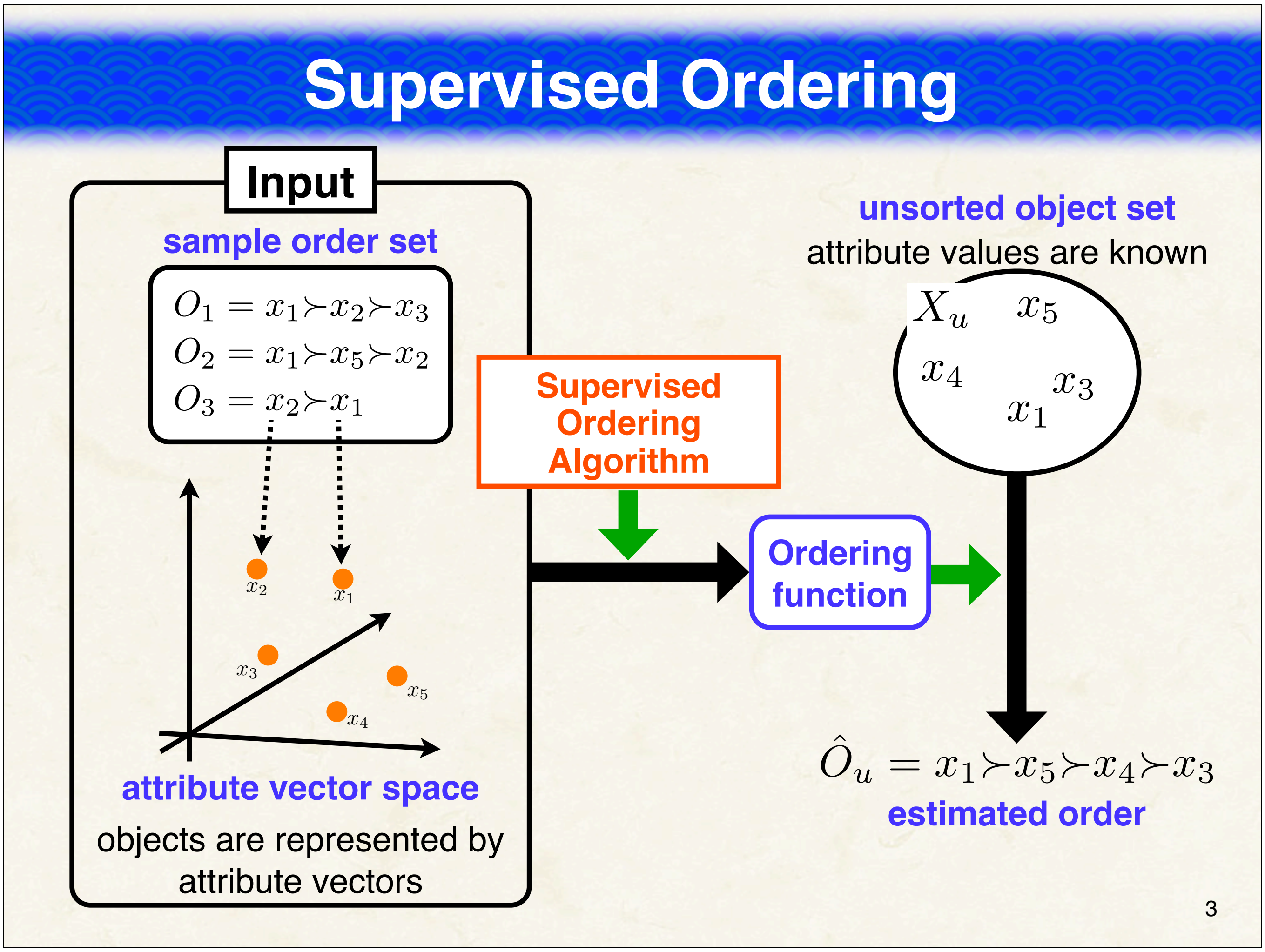
Toshihiro Kamishima and Shotaro Akaho
http://www.kamishima.net/
National Institute of Advanced Industrial Science and Technology (AIST), Japan

Int'l Conf. on Data Mining (ICDM2006) @ Hong Kong, China, 18-22/12/2006

Today I'd like to talk about a dimension reduction for supervised ordering .



A supervised ordering task is to learn a function for object ordering from given example orders.
For this task, the curse of dimensionality is serious problem like other learning tasks.
Therefore, we propose a dimension reduction technique specially designed for this task.
We begin with what is an order.
An order is an object sequence sorted according to a particular property.
For example, this is an order sorted according to my preference in sushi.
This order indicates that "I prefer a fatty tuna to squid", but "The degree of preference is unknown."



We first show an overview of a supervised ordering task.

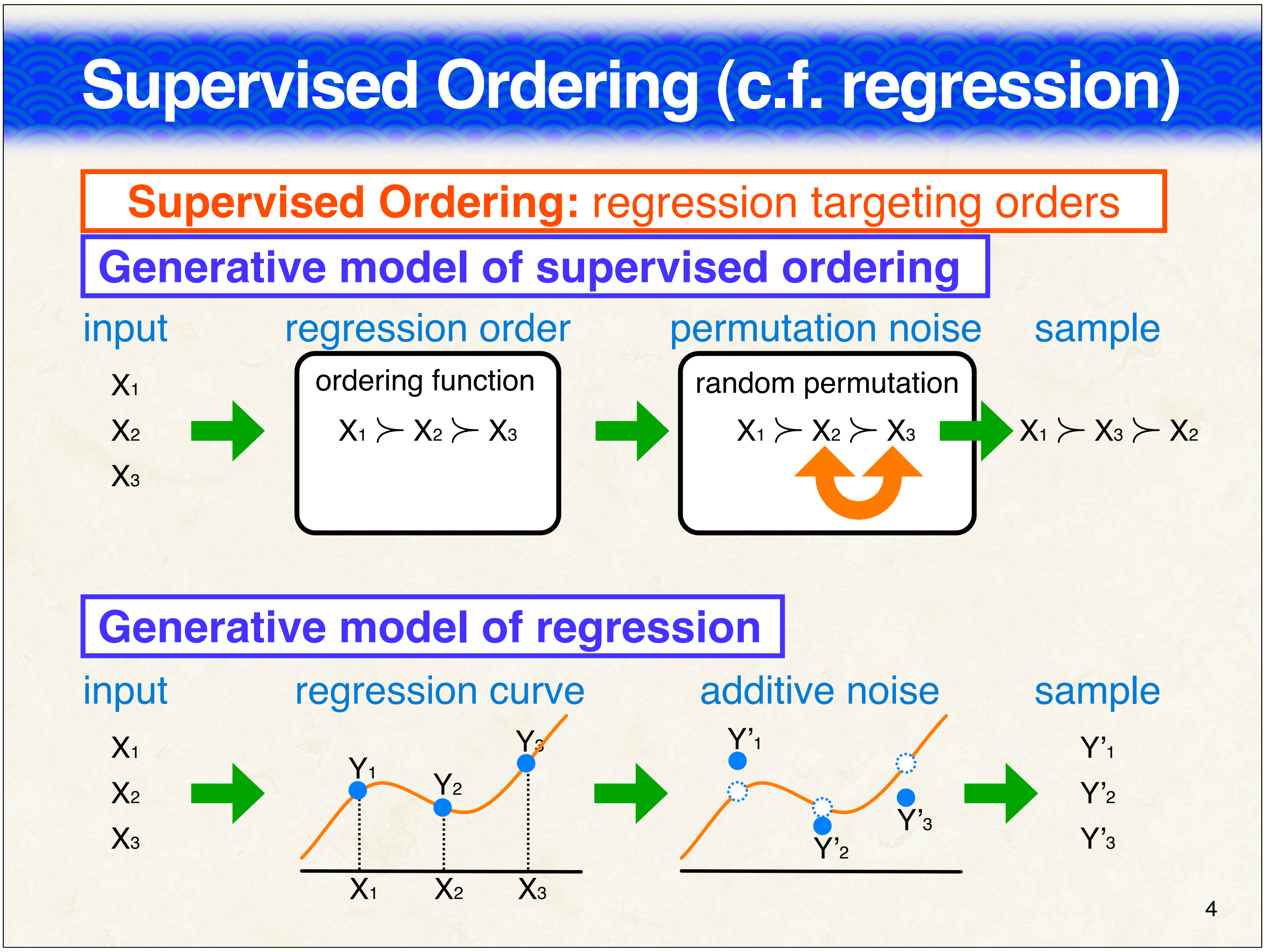
Training example orders are sorted according to the degree of the target property to learn.

Objects in these orders are represented by attribute vectors.

From these examples, a supervised ordering algorithm learn an ordering function.

By applying this learned function, unordered objects can be sorted according to the degree of the target property.

Objects not appeared in training examples can be ordered by referring the attribute values.



Supervised ordering can be considered as regression targeting orders.
This is a generative model of supervised ordering.
Unordered objects are given. These objects are sorted according to the degree of the target property.
This order is then affected by permutation noise, and finally a sample order is generated.
This model is very similar to that of regression, like this.

Ordinal Regression

regression whose dependent variables are ordered categorical

Supervised Ordering is more general problem than Ordinal Regression

Ordered Category can take one of predefined values, and, additionally, these values are ordered (ex. good-fair-poor)

Difference between "orders" and "ordered categories"

1. **Ordered Category** Absolute Information is contained
   Ex: for good-fair-poor, "good" indicates absolutely good.

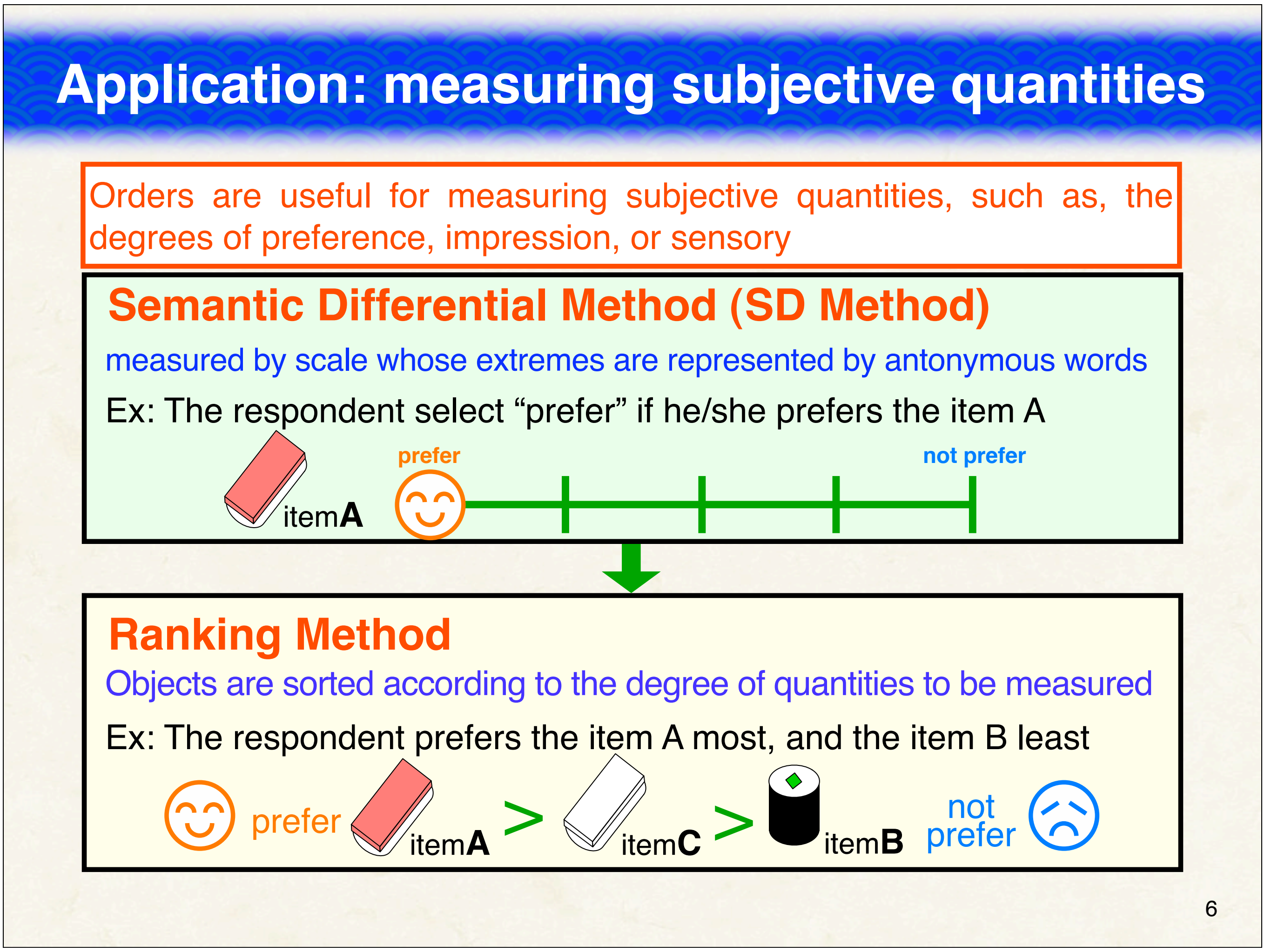
   Order It contains purely relative information

2. **Ordered Category** The # of grades is finite
   Ex: for good-fair-poor, the # of grades is limited to three

   Order The # of grades is not limited

5

This supervised ordering is also related to ordinal regression problem. The dependent variable of an ordinal regression task is ordered categorical variable.
Ordered categorical variables can take one of predefined values, and additionally, these values are ordered.
There are two points of differences between "orders" and "ordered category."
Ordered categorical values provide absolute information, and the number of grades is finite.
Therefore, supervised ordering is more general problem than ordinal regression.
Now, we have defined a supervised ordering task.
Next, we will show a few example tasks suited for using orders.



Orders are useful for measuring subjective quantities, such as the degrees of preference, impression, or sensory.

Such quantities can be measured by pointing on a scale like this.

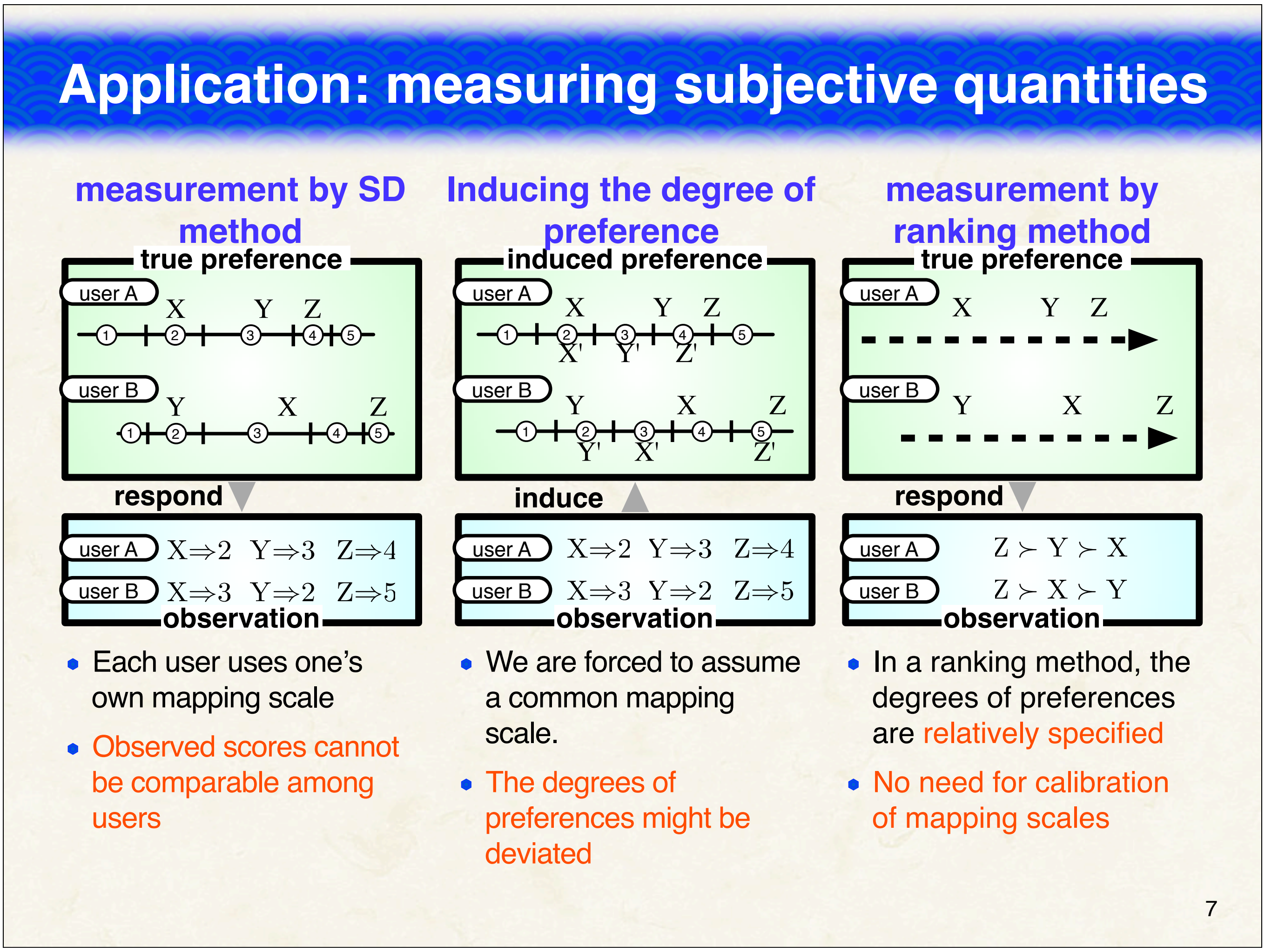
For example, the respondent select “prefer,” if he/she prefers the item A.

This is called an SD method.

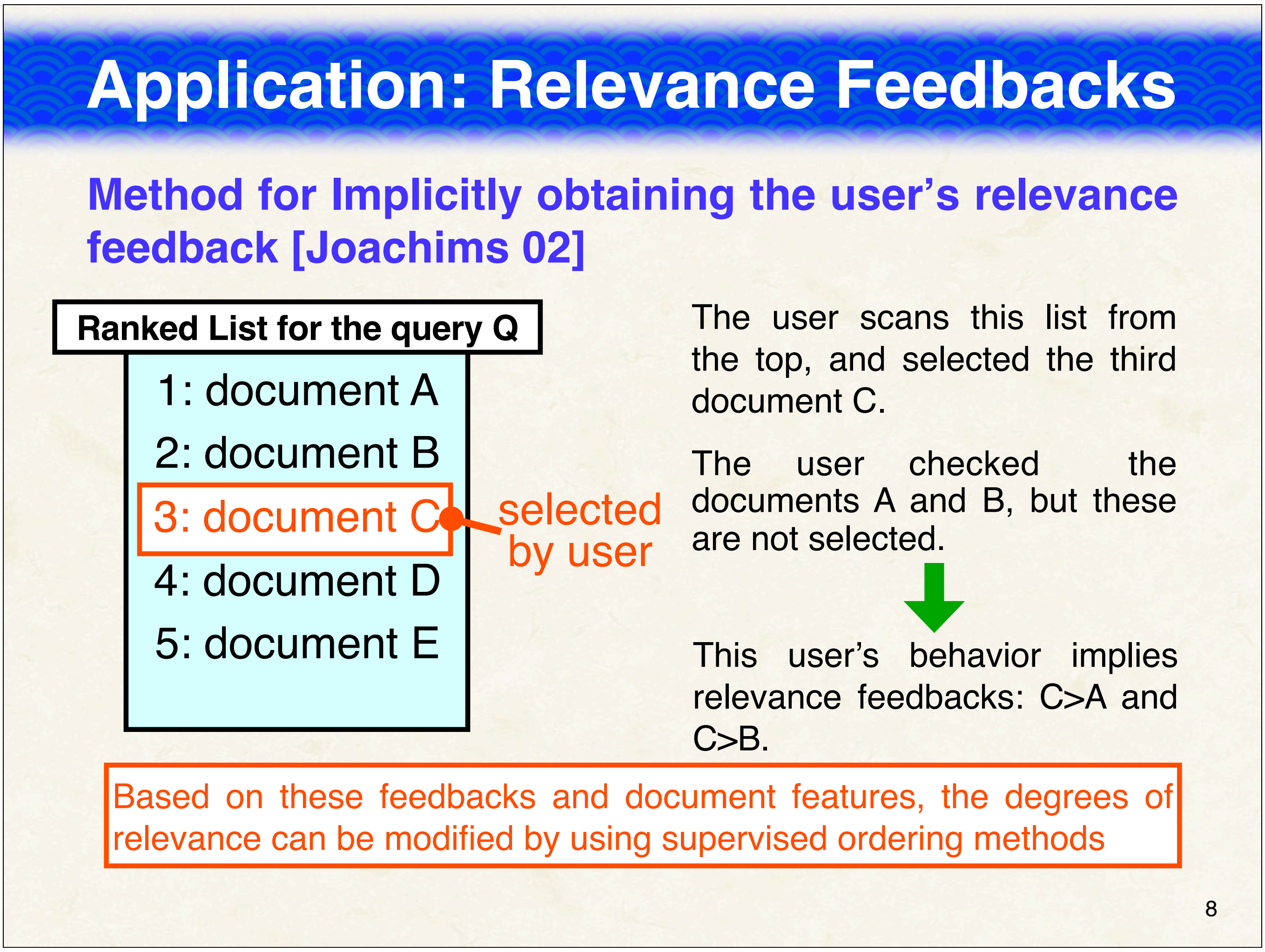
One alternative is a ranking method.

Objects are sorted according to the degree of quantities to be measured.

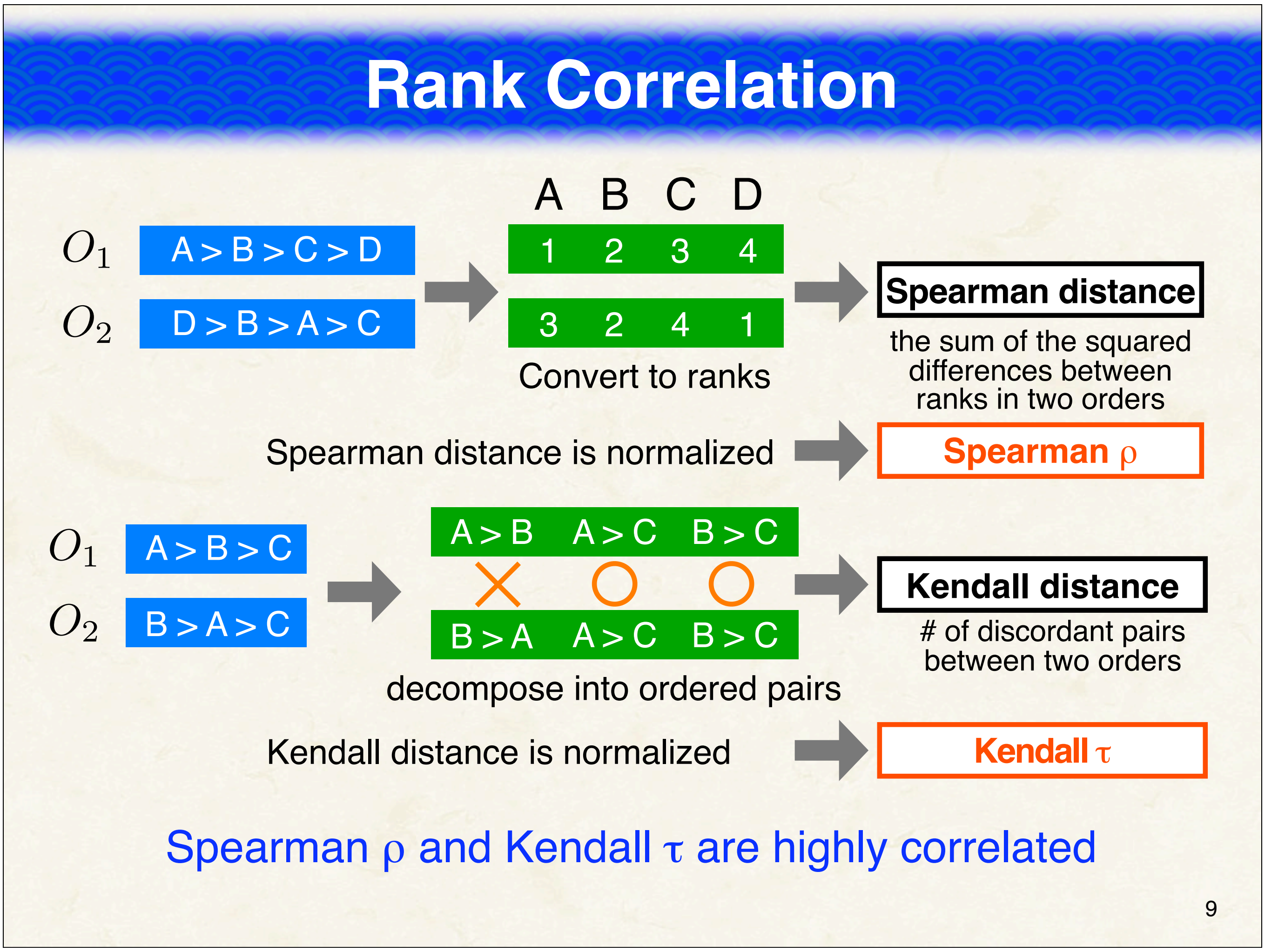
In this example, the user prefer Item A most, and the item B least.



We show a merit of using a ranking method.
We ask users their degree of preference, because the true degrees in users' mind cannot be observed directly.
For example, the degree of preference on the item X lies in interval 2 of user A; Then, the user A replies rating score 2.
Therefore, in an SD method, each user uses one's own mapping scale.
So, observed scores cannot be comparable among users.
Therefore, we are forced to assume a common mapping scale.
However, the degrees of quantities might be deviated to X to X'.
In a ranking method, the degrees of preferences are relatively specified.
So, no need for calibration of mapping scales.

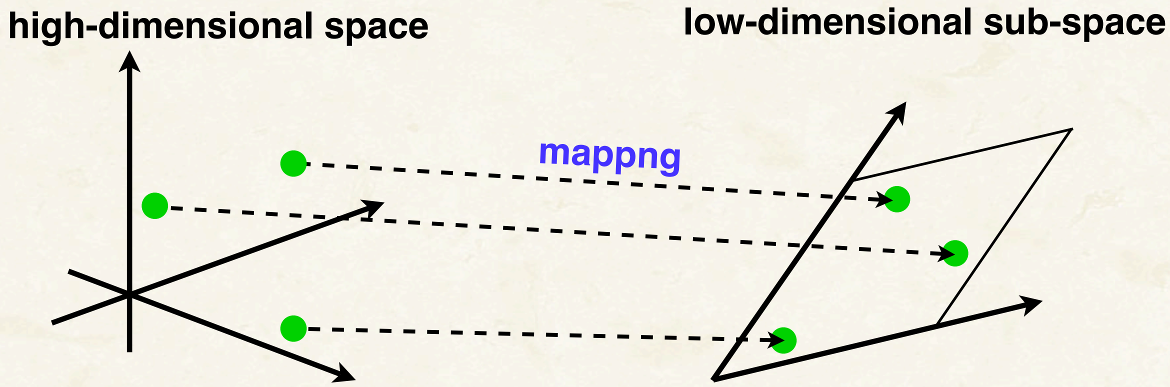


Orders are useful for dealing with relevance feedbacks.
Joachms proposed a method for implicitly obtaining the relevance feedbacks.
Given a ranked list for the query Q, the user scans this list from the top, and selected the third document C.
The user checked the documents A and B, but these are not selected.
This user's behavior implies relevance feedback: The document C is more relevant than the A or B.
Based on these feedbacks and document features, the degrees of relevance can be modified by using supervised ordering methods.
Now, we have shown usefulness of orders. Next, we will show a dimension reduction technique for a supervised ordering task.



Before showing our dimension reduction method, we show some basics.
Rank correlations, Spearman rho and Kendall tau, are widely used to measure the concordance between a pair of orders.
Spearman rho is calculated based on the sum of the squared differences of ranks in two orders.
Kendall tau is calculated by counting the number of discordant pairs between two orders.
These two rank correlations are highly correlated.

Dimension Reduction



mapping data in a high-dimensional space into a lower-dimensional sub-space, while limiting the amount of lost information

Ex. Principal Component Analysis, Fisher Discriminant Analysis

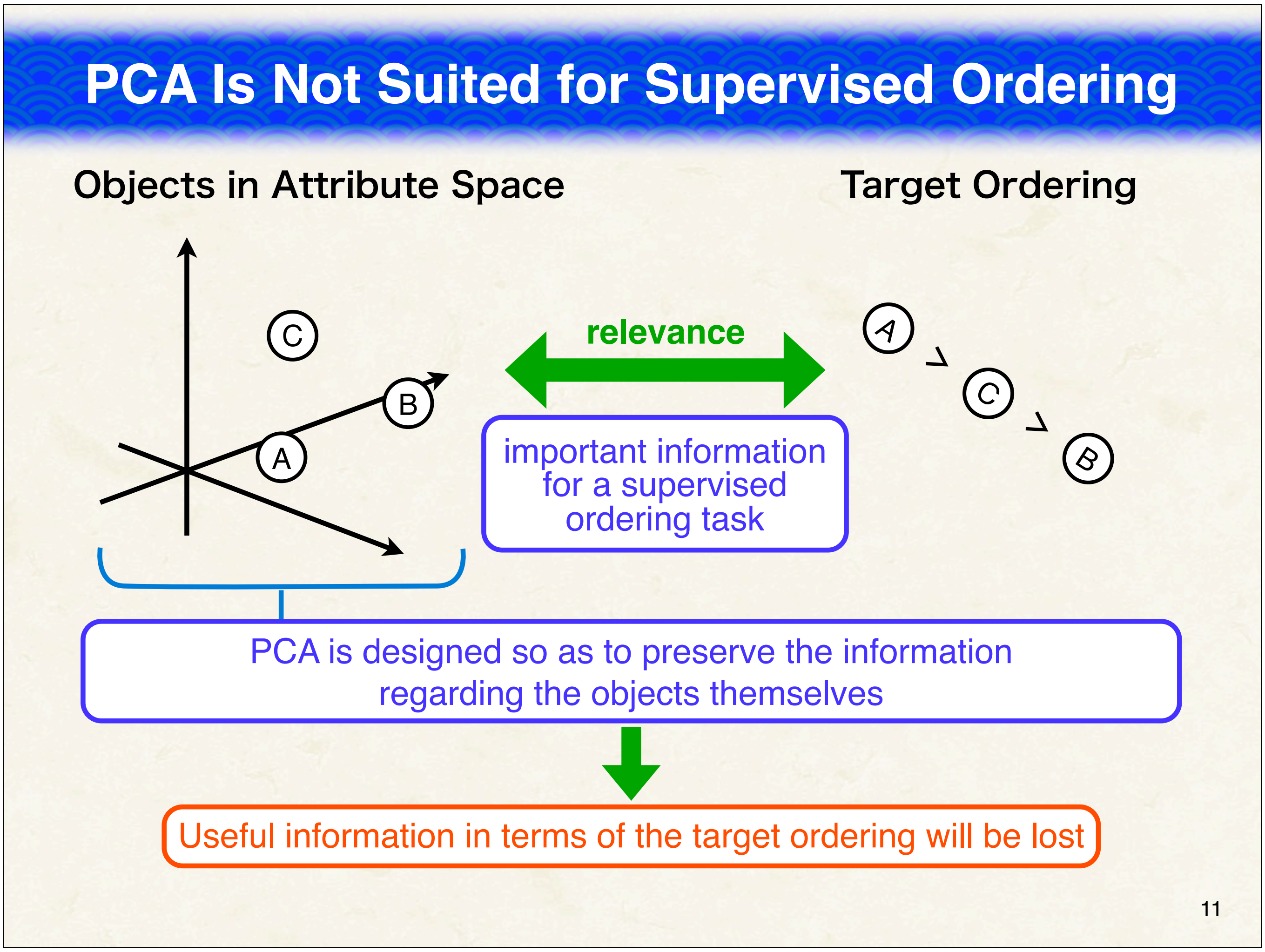
preprocess to avoid “the curse of dimensionality”

By finding the more informative sub-space, the model with high-generalization ability can be learned

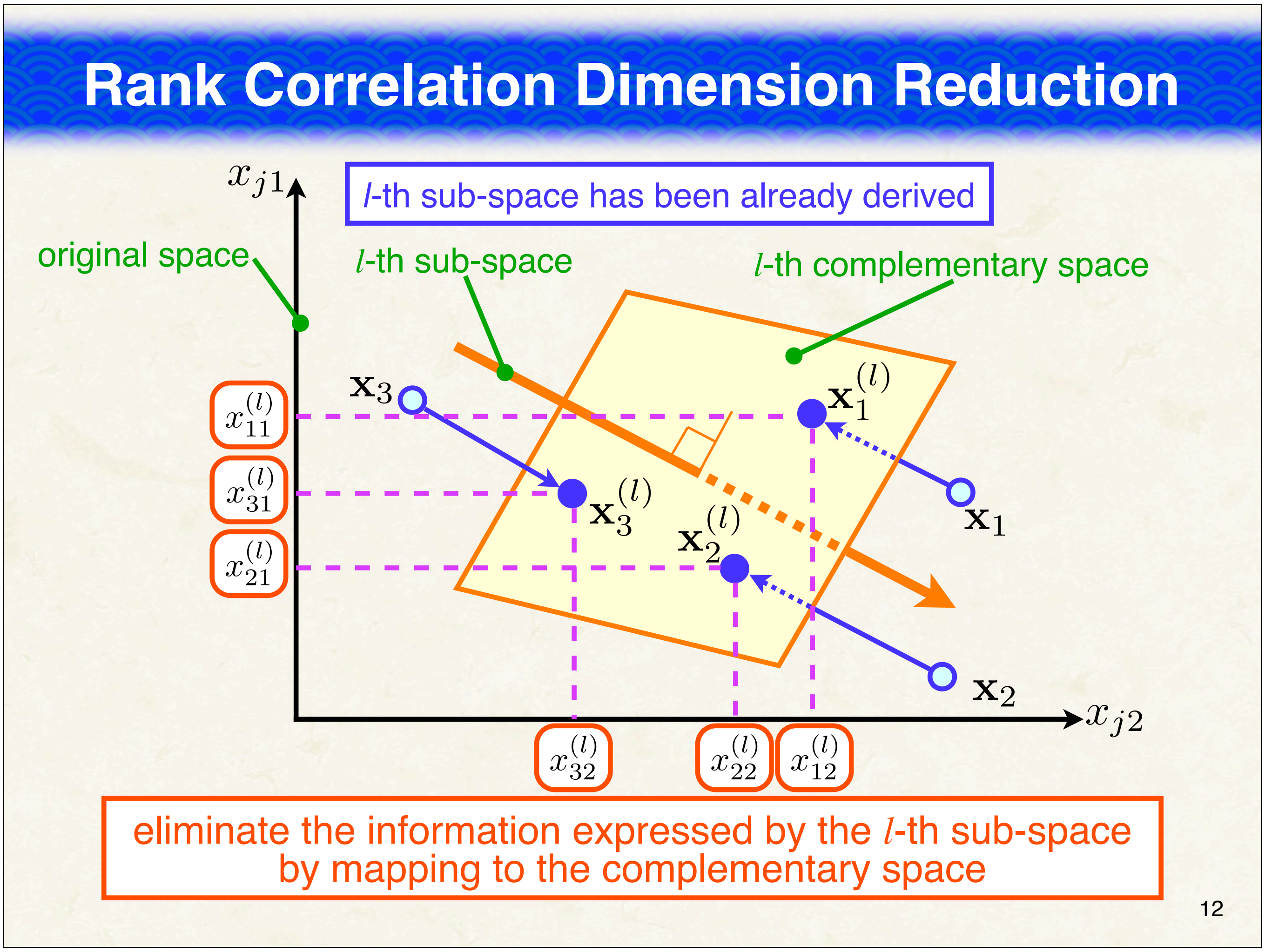
Dimension reduction is a technique for mapping data in a high-dimensional space into a lower-dimensional sub-space, while limiting the amount of lost information.

Dimension reduction is used as preprocess to avoid “the curse of dimensionality.”

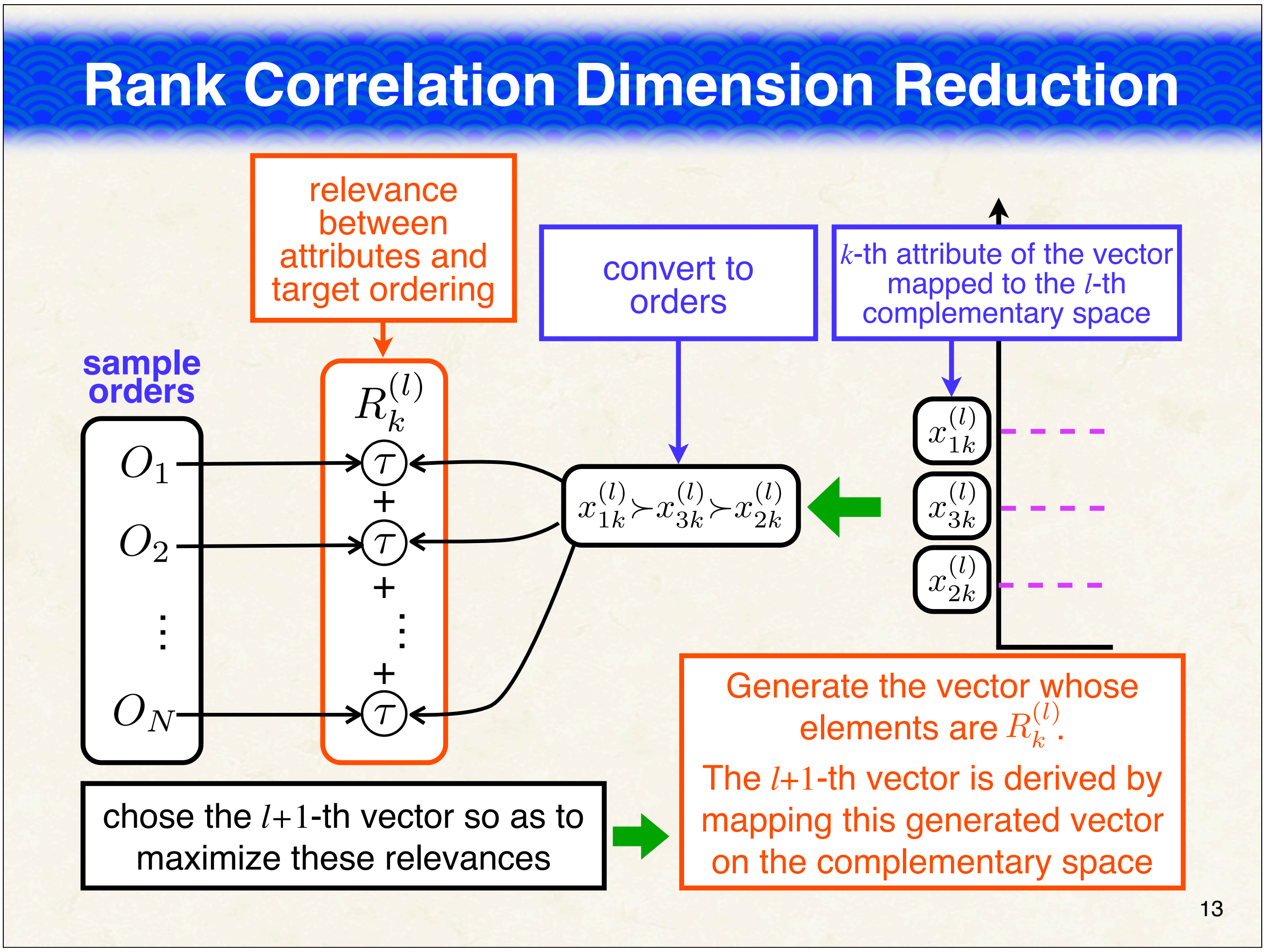
By finding the more informative sub-space, the model with high-generalization ability can be learned.



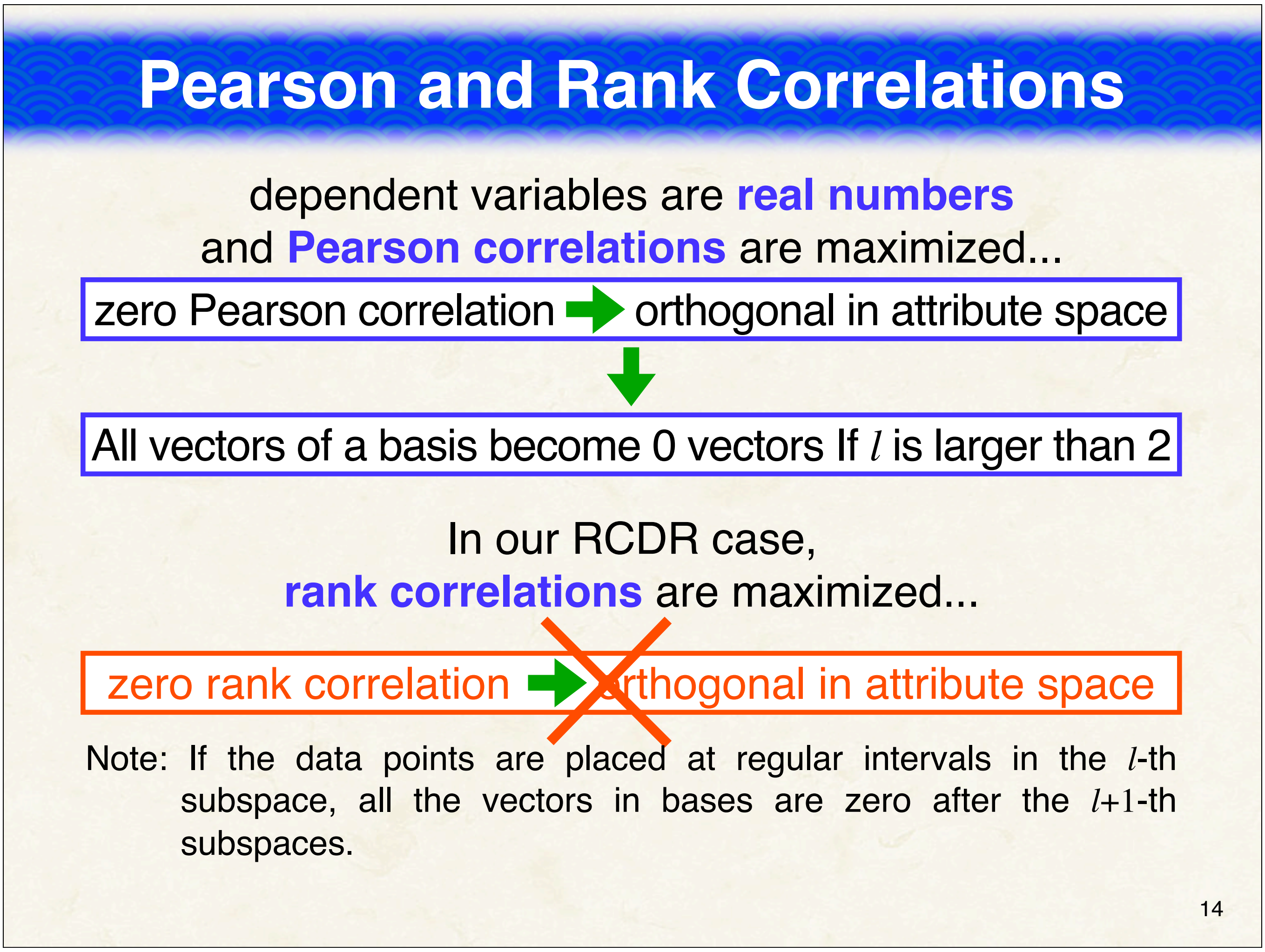
Principal component analysis is a widely used dimension reduction technique.
However, this PCA is not suited for solving a supervised ordering task.
PCA is designed so as to preserve the information regarding the objects themselves.
But, the important information for a supervised ordering task is the relevance between objects in attribute space and target ordering to learn.
Therefore, this useful information in terms of the target ordering will be lost.



Now, we will show our dimension reduction for a supervised ordering task.
In this method, vectors of basis are iteratively selected so as to preserve information about relevance between the attribute values and the target ordering.
We show one iteration process. We assume that the *l*-th sub-space has been already derived. In other words, we have *l* vectors of the basis, and try to find the next *l*+1-th vector.
First, we eliminate information expressed by this *l*-th sub-space.
To this aim, we consider the complementary space that is orthogonal to the l-th sub-space.
All the objects mapped to this complementary space.
After that, objects are represented by these mapped attribute vectors.



We observe the k-th attribute of the vector mapped to the l-th complementary space.
For solving a supervised ordering task, the ordinal information is more important. So, we convert to orders by sorting according to the mapped attribute values.
Next, we calculate the rank correlations between this converted order and for each sample order. These correlations are then summed up, and we get *Rk*. We consider this *Rk* represents the degree of relevance between attribute values in the l-th complementary space and the target ordering.
Now, all that we have to do is to chose the l+1-th vector of the basis so as to maximize these relevances in the l-th complementary space.
To this aim, we calculate this Rk for each attribute, and generate the vector whose elements are Rk. The l+1-th vector of the basis is derived by mapping this generated vector on the l-th complementary space.
By iterating these process, we can obtain the sub-space. We call this method, rank correlation dimension reduction.



Here, we want to insist the difference between Pearson and rank correlations.
Assume that dependent variables are real numbers and Pearson correlation is maximized.
In this case, zero Pearson correlation implies the orthogonality in attribute space.
Therefore, all vectors of a basis are zero vectors, if l is larger than 2.
On the other hand, in our RCDR case, rank correlations are maximized.
In this case, zero rank correlation doesn't implies orthogonality.
Therefore, vectors of a basis can be non-zero vectors, even if l is larger than 2.
Now, we have shown our new RCDR method. Next, we will show simple example and experimental results.

Simple Example (1)

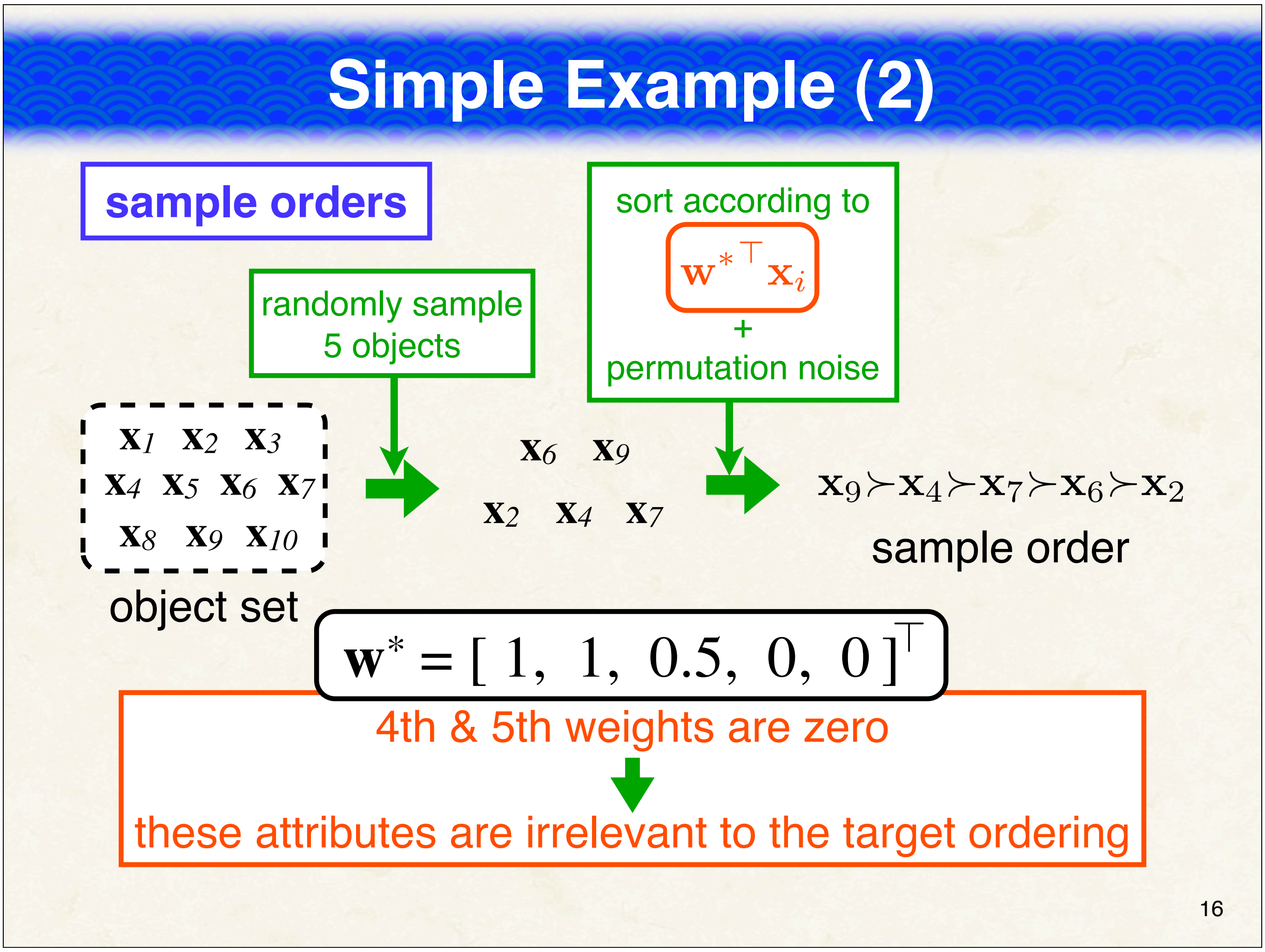
original attribute vector

- each object is represented by 5 attributes
- 1000 objects are randomly generated
- each attribute values follow normal distribution N(0,1)
- correlations among attributes are designed ...

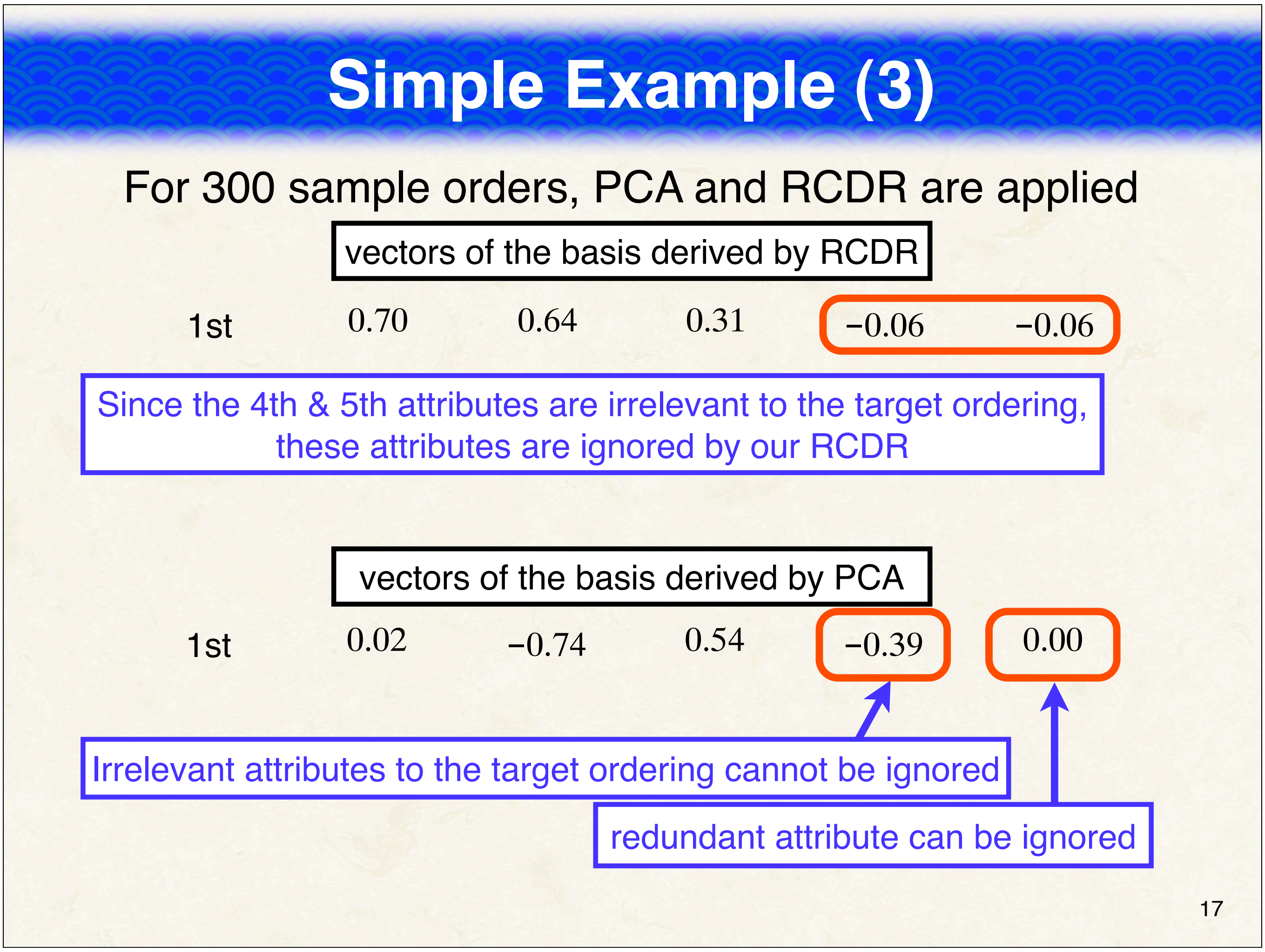
perfectly corrleated
= completely equal

$$\mathbf{x}_i = [\underbrace{x_{i1},\ x_{i2},\ x_{i3},\ x_{i4}}_{\text{mutually independent}},\ x_{i5}]^\top$$

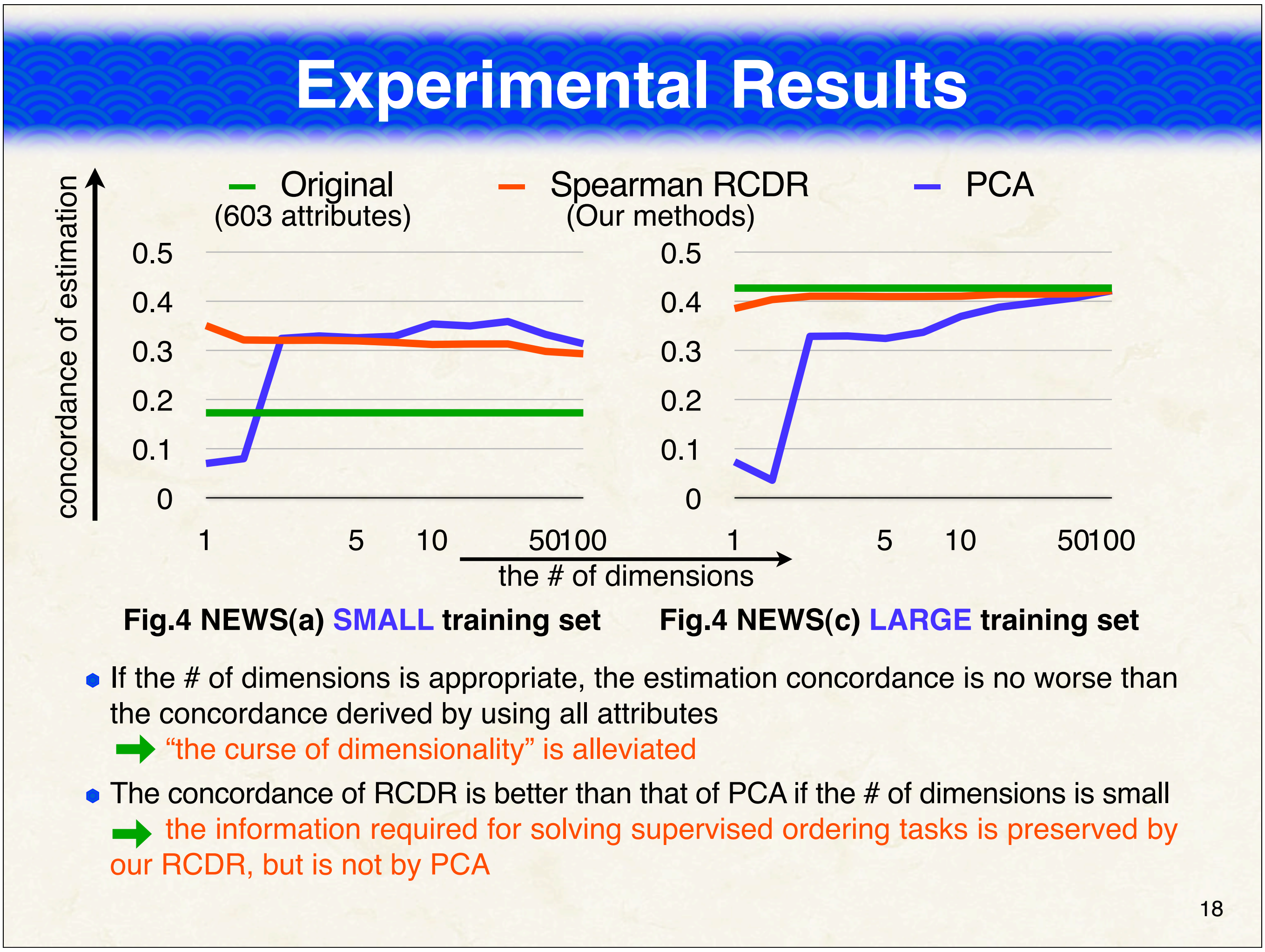
We first show a simple example to demonstrate what is produced by our RCDR methods.
Each object is represented by 5 attributes.
1000 objects are randomly generated.
Each attribute value follows normal distribution with zero mean and unit variance.
Correlations among features are designed like this.
The first to fourth attributes are mutually independent. The fourth and fifth attributes are perfectly correlated. That is to say these two attributes are completely equal.
In this case, by applying a PCA technique, one of these two attributes will be considered redundant, and will be ignored.



From these generated objects, sample orders are constructed.
First, we randomly sample five objects.
These objects are sorted according to the weighted sum of attribute values.
Then, permutation noise is added to this order. Finally, a sample order is obtained.
This process is repeated.
Here, we use weights of attributes like this.
Because the forth and fifth weights are zero, these attributes are irrelevant to the target ordering.
Therefore, these two attributes will be ignored by applying our RCDR method.



For 300 sample orders, PCA and RCDR techniques are applied.
In the first vector of the basis derived by our RCDR, these two elements are nearly zero.
Since the fourth and fifth attributes are irrelevant to the target ordering, these attributes are ignored by our RCDR.
In the case of PCA, the fifth attribute is zero, because redundant attribute can be ignored by applying PCA.
However, the fourth element is not zero, because irrelevant attributes to the target ordering cannot be ignored.



Next, we show experimental results on real data.
News articles are sorted by users according to their significance. Based on word and category attributes, these orders are estimated by using supervised ordering techniques.
In these charts, the upper indicates the better estimation. Green lines show results derived by using all original 603 attributes. Red and Blue lines show results derived after preprocessed by our RCDR and PCA, respectively. The number of dimensions is varied.
If the number of dimensions is appropriate, the estimation concordance is no worse than the concordance derived based on all attributes.
This fact indicates that the curse of dimensionality is alleviated.
This is because the simpler and more appropriate model class can be used in a learning process.
The concordance of our RCDR is better than that of PCA if the number of attributes is small
This fact indicates that the information required for solving supervised ordering tasks is preserved by our RCDR, but is not by PCA

Conclusion

- **Rank Correlation Dimension Reduction**
  To performing supervised ordering tasks, information about relation between target ordering and features of objects. Therefore, this reduction technique is specially designed so as to preserve this information.
  - In a case of supervised ordering task, estimation concordance derived by using our RCDR is superior to that derived by general purpose PCA.
  - If the # of training samples is small, generalization ability can be improved by using RCDR techniques.
  - Performances of two rank correlations, Kendall τ and Spearman ρ, are almost equivalent.

more infomation: `http://www.kamishima.net/`

We would like to conclude our talk.
Our contributions are as follows.
More information is available in our Web site.
That's all we have to say. Thank you for your attension.

Bibliography

- [Cohen 99] W.W.Cohen, R.E.Schapire, Y.Singer. Learning to Order Things. Journal of Artificial Intelligence Research, Vol.10, pp.243-270 (1999)
- [Freund 98] Y.Freund, R.Iyer, R.E.Schapire, and Y.Singer. An Efficient Boosting Algorithm for Combining Preferences. In Proc. of The 15th Int'l Conf. on Machine Learning, pp.170-178 (1998)
- [Herbrich 98] R.Herbrich, T.Graepel, P.Bollmann-Sdorra, K.Obermayer. Learning Preference Relations for Information Retrieval. In: ICML-98 Workshop: Text Categorization and Machine Learning, pp.80-84, (1998)
- [Joachims 02] T.Joachims. Optimizing Search Engines Using Clickthrough Data. In Proc. of The 8th Int'l Conf. on Knowledge Discovery and Data Mining, pp.133-142 (2002)
- [Kamishima 05] T.Kamishima, H.Kazawa, and S.Akaho. Supervised Ordering — An Empirical Survey. In Proc. of The 5th IEEE Int'l Conf. on Data Mining (2005)
- [Kazawa 03] H.Kazawa, T.Hirao, and E.Maeda. Order SVM: a keernel method for order learning based on generalized order statistics. Systems and Computers in Japan, vol.36, no.1, pp.35-43 (2005)