Advance Machine Learning

01CO1301
4 Credits

Marwadi University

Department of Computer Engineering

Ravikumar R Natarajan

Course Outcomes

- At the end of the course, students will be able to:
- To understand key concepts, tools and approaches for pattern recognition on complex data sets.
- To learn Kernel methods for handling high dimensional and non-linear patterns.
- To implement state-of-the-art algorithms such as Support Vector Machines and Bayesian networks.
- To Solve real-world machine learning tasks: from data to inference.
- To apply theoretical concepts and the motivations behind different learning frameworks.

Dimensionality Reduction

Unit #4

Marwadi University

Department of Computer Engineering

Content

- Dimensionality Reduction
  - CCA
  - LDA
  - ICA
  - NMF
- Canonical Variates
  - Feature Selection vs Feature Extraction

CCA

- Canonical correlation analysis (CCA) is a statistical technique that identifies a sequence of pairs of **patterns in two multivariate data** sets and constructs sets of transformed variables by projecting the original data onto these patterns.

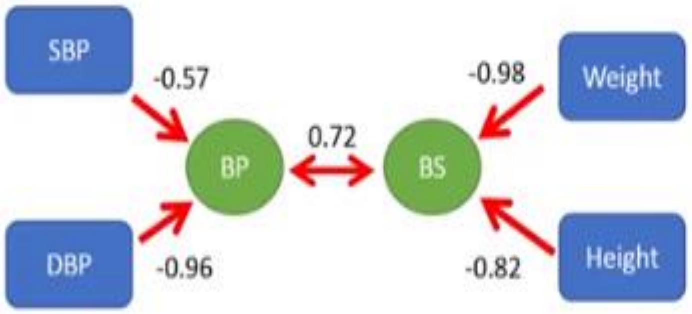
- Canonical correlation analysis is used to identify and measure the associations among two sets of variables. Canonical correlation is appropriate in the same situations where multiple regression would be, but where are there are multiple inter-correlated outcome variables.

- from sklearn.cross_decomposition.CCA

CCA

	SBP	DBP	Height	Weight
1	120	76	165	60
2	109	80	180	80
3	130	82	170	70
4	121	78	185	85
5	135	85	180	90
6	140	87	187	87

BP BS

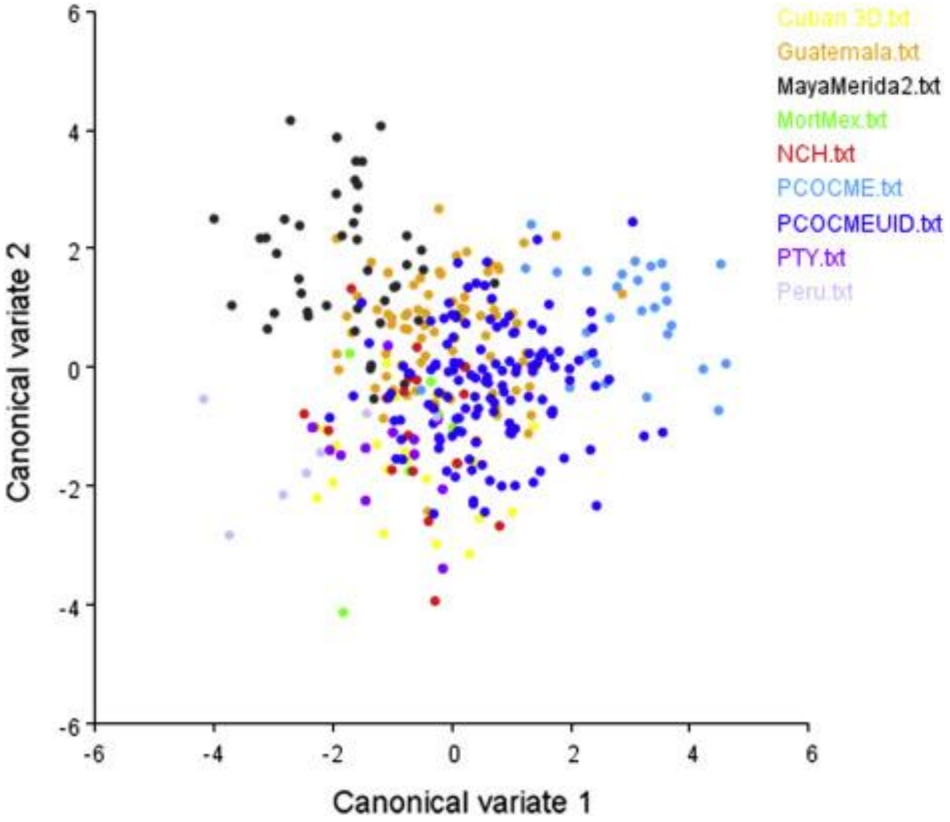


Correlation matrix

	SBP	DBP	Height	Weight
SBP	1			
DBP	0.79	1		
Height	0.25	0.54	1	
Weight	0.37	0.66	0.92	1

Canonical Variates

- The canonical variates are linear combinations of the original variables that maximally separate groups.
- Similar to PCA, individual or group mean scores can be plotted to interpret patterns of variation across the groups being analyzed.



LDA

- Both PCA and LDA are linear transformation techniques. However, PCA is an unsupervised while LDA is a supervised dimensionality reduction technique.

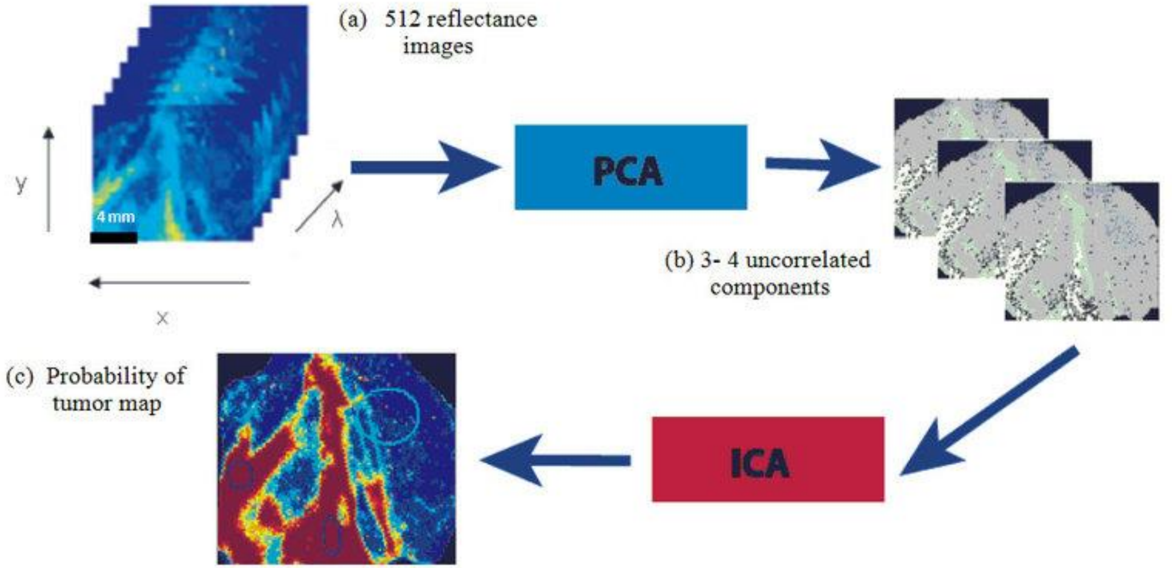
- LDA (Linear Discriminant Analysis) tries to reduce dimensions of the feature set while retaining the information that discriminates output classes. LDA tries to find a decision boundary around each cluster of a class. It then projects the data points to new dimensions in a way that the clusters are as separate from each other as possible and the individual elements within a cluster are as close to the centroid of the cluster as possible.

- The new dimensions are ranked on the basis of their ability to maximize the distance between the clusters and minimize the distance between the data points within a cluster and their centroids. These new dimensions form the linear discriminants of the feature set.

ICA

- PCA is often used to compress information i.e. dimensionality reduction. While ICA (Independent Component Analysis) aims to separate information by transforming the input space into a maximally independent basis.

- ICA is an important tool in neuroimaging, MRI, and EEG analysis that helps in separating normal signals from abnormal ones.



PCA looks for uncorrelated factors while ICA looks for independent factors.

NMF

- Nonnegative matrix factorization NMF is a linear powerful technique for dimension reduction.
- It reduces the dimensions of data making learning algorithms faster and more effective.
- This is a very strong algorithm which many applications. For example, it can be applied for Recommender Systems, for Collaborative Filtering for topic modelling and for dimensionality reduction.
- In Python, it can work with sparse matrix where the only restriction is that the values should be non-negative.

DEMO

- https://colab.research.google.com/drive/1ohpNqMtBH9uAiGiQxuSRSxzbmmZf08-u?usp=sharing

- Kaggle Reference
- https://www.kaggle.com/code/ljlbarbosa/dimensionality-reduction-in-python/notebook

Feature Selection vs Feature Extraction

Feature selection is about selecting a subset of features out of the original features in order to reduce model complexity, enhance the computational efficiency of the models and reduce generalization error introduced due to noise by irrelevant features. The following represents some of the important feature selection techniques:

- **Regularization techniques**
- **Feature importance** techniques such as using estimator such as Random Forest algorithm to fit a model and select features based on the value of attribute such as feature_importances_
- **Greedy search algorithms** such as some of the following which are useful for algorithms (such as K-nearest neighbours, K-NN) where regularization techniques are not supported.
- **Filter methods**: The filter model only considers the association between the feature and the class label
- **Wrapper methods**
- **Embedded methods**: In embedded method, the features are selected in the training process of learning model, and the feature selection result outputs automatically while the training process is finished. Training the Lasso regression model is a classic example of embedded method for feature selection.

Feature Selection vs Feature Extraction

Feature extraction is about extracting/deriving information from the original features set to create a new features subspace. The primary idea behind feature extraction is to compress the data with the goal of maintaining most of the relevant information.

- The following are different types of feature extraction techniques:
- **PCA**
- **LDA**

End of Unit 4

Thank You.