Syllabus for the Datamining Class

Gener Avilés R

2017-02-22

Contents

Chapter 1

Introduction

Chapter 2

Principal Components Analysis

Chapter 3

Canonical Correlation Analysis (CCA)

3.1 What is CCA?

- Seeks the weighted linear composit for each variate (sets of D.V. or I.V.) to maximize the overlap in their distributions.
- Labeling of DV and IV is arbitrary. The procedure looks for relationships and not causation.
- Goal is to **maximize the correlation** (not the variance extracted as in most other techniques).
- Lacks specificity in interpreting results, that may limit its usefulness in many situations.

CCA helps us answer the question: ***What is the best way to understand how the variable sin two sets are related?*** , mathematically speaking: ***what linear combinations of the*** X ***variables (***u***) and the Y vairables (***t***) will maximize their correlation?***

This chapter is under construction.

Figure 3.1:

Bibliography