Canonical Correlation Analysis (CCA)
HessMCC
Multiview Dimension Reduction Via Hessian Multiset Canonical Correalation

Mamta Pawar (0801CS183D10)

December 13, 2020

Contents

Chapter 1
Introduction

1.1 Multiview Learning

Multi-view learning (MVL) is a strategy for fusing data from different sources or subsets. Canonical correlation analysis (CCA) is very important in MVL, whose main idea is to map data from different views onto a common space with the maximum correlation. The traditional CCA can only be used to calculate the linear correlation between two views. Moreover, it is unsupervised, and the label information is wasted in supervised learning tasks. Many nonlinear, supervised, or generalized extensions have been proposed to overcome these limitations. However, to our knowledge, there is no up-to-date overview of these approaches. Integrate Hessian into the multiset canonical correlations and derive Hessian multiset canonical correlations (HesMCC). HesMCC takes the advantage of Hessian and provides superior extrapolating capability. Therefore, HesMCC can significantly leverage the performance.

Hessian regularization into multiset CCA for multiview dimension reduction

Hessian multiset canonical correlations (HesMCC) for multiview dimension reduction. Hessian can properly exploit the intrinsic local geometry of the data manifold in contrast to Laplacian. HesMCC takes the advantage of Hessian and provides superior extrapolating capability and finally leverage the performance

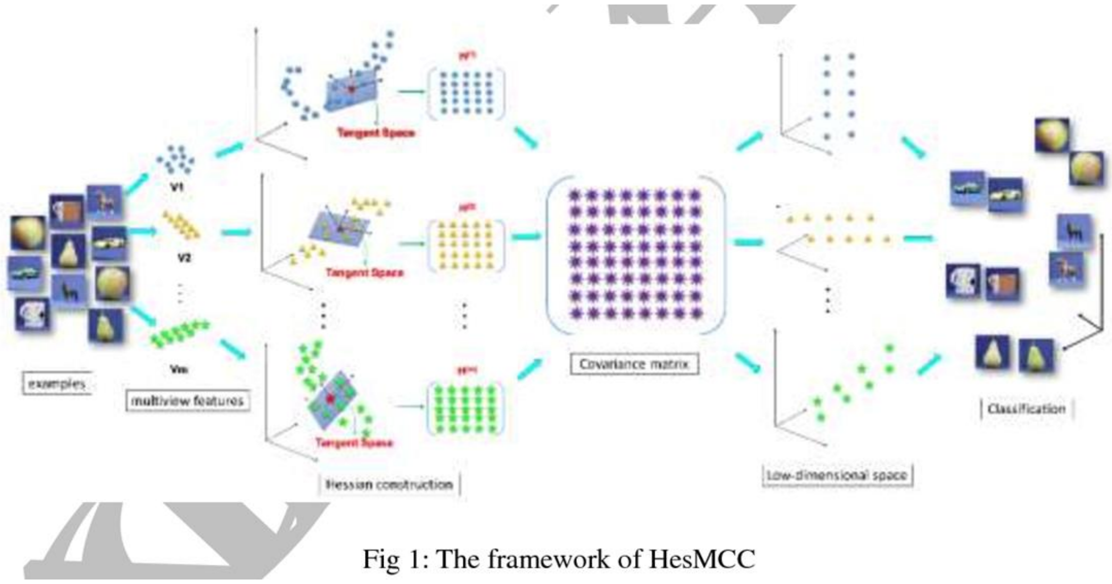


Fig 1: The framework of HesMCC

1.2 CCA

Canonical correlation analysis (CCA) is a way of measuring the linear relationship between two multi-dimensional variables. It finds two bases, one for each variable, that are optimal with respect to correlations and, at the same time, it finds the corresponding correlations.

For example, an image can be described with color, shape, and texture features. The extracted visual features usually have high dimensions of up to hundreds or thousands, which often causes the problem called the curse-of-dimensionality. Hence multiview dimension reduction algorithms have been subsequently proposed with the purpose of finding an appropriate low-dimensional feature subspace from multiview high dimensional features. Canonical correlation analysis (CCA) is one of the most representative techniques and has been widely applied to many multiview learning applications including classification, retrieval, regression and clustering Canonical correlation analysis (CCA) proposed by Hotelling seeks a pair of linear transformation for two view high dimensional features such that the corresponding low-dimensional projections are maximally correlated.

Chapter 2

Mathematical Formulation

Hessian Matrix:
Hessian matrix or **Hessian** is a square **matrix** of second-order partial derivatives of a scalar-valued function, or scalar field. It describes the local curvature of a function of many variables.

$$H\,f\,(x_1, x_2, \ldots, x_n) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\,\partial x_2} & \frac{\partial^2 f}{\partial x_1\,\partial x_3} & \cdots & \frac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \frac{\partial^2 f}{\partial x_2\,\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \frac{\partial^2 f}{\partial x_2\,\partial x_3} & \cdots & \frac{\partial^2 f}{\partial x_2\,\partial x_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n\,\partial x_1} & \frac{\partial^2 f}{\partial x_n\,\partial x_2} & \frac{\partial^2 f}{\partial x_n\,\partial x_3} & \cdots & \frac{\partial^2 f}{\partial x_n^2}. \end{bmatrix}$$

$$H_x = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix}$$

For a function $f: \mathbb{R}^d \to \mathbb{R}$, its Hessian matrix at $\boldsymbol{x} = (x_1, \ldots, x_d)^\top$ is defined as:

$$\boldsymbol{H} = \begin{bmatrix} \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1^2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_d \partial x_1} \\ \vdots & & \vdots \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_d} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_d^2} \end{bmatrix}.$$

Use the notation in the matrix calculus handout and prove that

$$\boldsymbol{H} = \frac{\partial}{\partial \boldsymbol{x}} \left(\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} \right)^\top.$$

Chapter 3

CCA Algorithm

3.1HessCCA Algorithm

Second Derivative Test, the general n variable version:
Suppose that the second partial derivatives of $f : \mathbb{R}^n \to \mathbb{R}$ are continuous on a ball with centre $\vec{c}$, where $\nabla f(\vec{c}) = \vec{0}$ (that is, $\vec{c}$ is a critical point of f).

Let H denote the Hessian matrix of second partial derivatives, and for each $k = 1, 2, ..., n$, let D_k denote the determinant of the Hessian in the variables x_1, x_2, ..., x_k. Assume that $|H(\vec{c})| \neq 0$.

(a) If $D_k(\vec{c}) > 0$ for all $k = 1, 2, ..., n$, then f has a local minimum at $\vec{c}$.
(b) $(-1)^k \cdot D_k(\vec{c}) > 0$ for all $k = 1, 2, ..., n$, then f has a local maximum at $\vec{c}$.
(c) Otherwise, f has a saddle point at $\vec{c}$. ■

This theorem is usually proved using the quadratic approximation of the (multivariable) Taylor Series for f centred at $\vec{x} = \vec{c}$ and understandably involves a good amount of Linear Algebra. The curious student may consult an Advanced Calculus textbook for a proof of this theorem.

Chapter 4

Documentation of API

4.1 Package organization

```
from sympy import *
import math
```

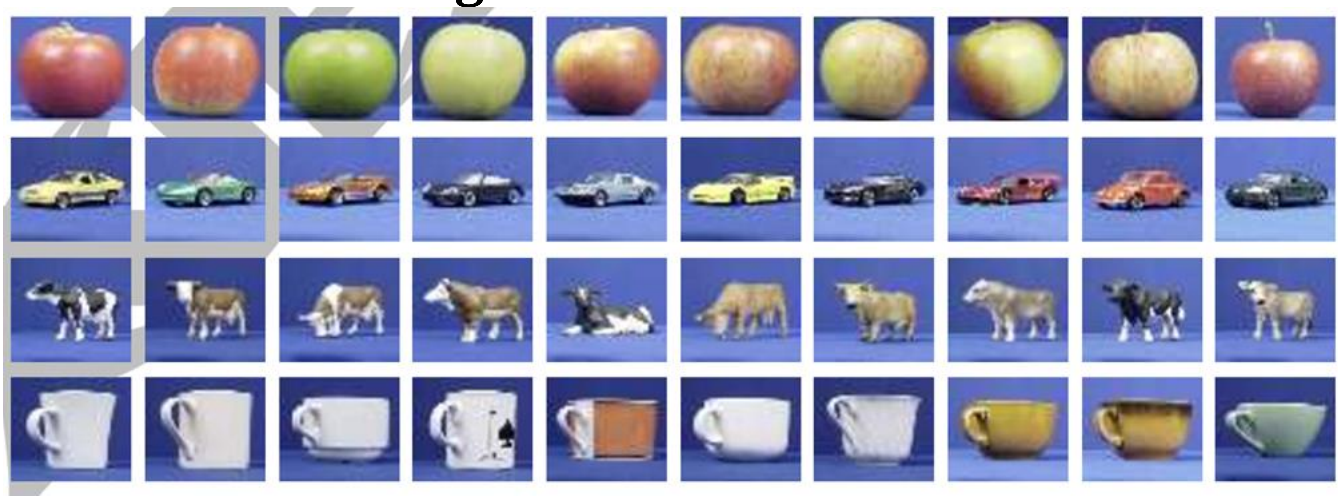
4.2 Methods

diff()

Chapter 5

Example

5.1 Multiview images



Chapter 6

Learning Outcomes

- Steps of optimization using Hessian Matrix.
- Determine the roles of derivatives in optimization using hessian matrix.
- Significance hessian matrix in optimization used in case of multiple input dimension.

Chapter 7

References

https://math.stackexchange.com/questions/2324806/how-can-i-prove-that-hessian-of-the-function-fx-1-x-6-is-degenerate

https://www.tandfonline.com/doi/full/10.1080/21642583.2018.1545610