Abteilung Maschinelles Lernen
Institut für Softwaretechnik und theoretische Informatik
Fakultät IV, Technische Universität Berlin
Prof. Dr. Klaus-Robert Müller
Email: klaus-robert.mueller@tu-berlin.de

Machine Learning 1

Winter semester 2016/17

Group APXNLE

Exercise 5

Members

Jing Li[387272] jing.li.1@campus.tu-berlin.de

Kumar Awanish[386697] k.awanish@campus.tu-berlin.de

Manjiao Xu[386498] manjiao.xu@campus.tu-berlin.de

Rudresha Gulaganjihalli Parameshappa[386642]

Gulaganjihalliparameshappa@ campus.tu-berlin.de

Sonali Nayak[386995] sonali.nayak@campus.tu-berlin.de

Maximilian Ernst[364862] maximilian.ernst@campus.tu-berlin.de

Exercise 1: Finding the direction of maximal correlation between datasets

(a) We apply linear projection to x and y.

$w_x^T X \ , \ w_y^T Y$

$Var(X) = \Sigma_{XX} \qquad Var(Y) = \Sigma_{YY}$

$$\rho = Corr(\vec{w}_x^T X, \vec{w}_y^T Y) = Cov(\vec{w}_x^T X, \vec{w}_y^T Y) / var(\vec{w}_x^T, X)^{\frac{1}{2}} \cdot var(\vec{w}_y^T, Y)^{\frac{1}{2}}$$

$$= \vec{w}_x^T cov(X, Y) \vec{w}_y / (\vec{w}_x^T Var(X) \vec{w}_x)^{\frac{1}{2}} \cdot (\vec{w}_y^T Var(Y) \vec{w}_y)^{\frac{1}{2}}$$

$$= \frac{\vec{w}_x^T \Sigma_{XY} \vec{w}_y}{(\vec{w}_x^T \Sigma_{XX} \vec{w}_x)^{\frac{1}{2}} (\vec{w}_y^T \Sigma_{YY} \vec{w}_y)^{\frac{1}{2}}}$$

The optimization problem is

$$\max_{\vec{w}_x, \vec{w}_y} \frac{\vec{w}_x^T \Sigma_{XY} \vec{w}_y}{(\vec{w}_x^T \Sigma_{XX} \vec{w}_x)^{\frac{1}{2}} (\vec{w}_y \Sigma_{YY} \vec{w}_y)^{\frac{1}{2}}}$$

(b) Let $\vec{u} = \Sigma_{XX}^{\frac{1}{2}} \vec{w}_x \qquad \vec{v} = \Sigma_{YY}^{\frac{1}{2}} \vec{w}_y$

$$\rho = \vec{u}^T \Sigma_{XX}^{\frac{1}{2}} \Sigma_{XY} \Sigma_{YY}^{\frac{1}{2}} \vec{v} / (\vec{u}^T \vec{u})^{\frac{1}{2}} \cdot (\vec{v}^T v)^{\frac{1}{2}}$$

According to Cauchy-Schwarz inequality, we have

$$\vec{u}^T \Sigma_{XX}^{\frac{1}{2}} \Sigma_{XY} \Sigma_{XY}^{\frac{1}{2}} \vec{v} \le (\vec{u}^T \Sigma_{XX}^{\frac{1}{2}} \Sigma_{XY} \Sigma_{YY}^{\frac{1}{2}} \Sigma_{YY}^{\frac{1}{2}} \Sigma_{YX} \Sigma_{XX}^{\frac{1}{2}} \vec{u})^{\frac{1}{2}} \cdot (\vec{v}^T v)^{\frac{1}{2}}$$

$$\rho = Corr(\vec{w}_x^T X, \vec{w}_Y^T Y) \le \frac{(\vec{u}^T \Sigma_{XX}^{\frac{1}{2}} \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{YX} \Sigma_{XX}^{\frac{1}{2}} \vec{u})^{\frac{1}{2}}}{(\vec{v}^T v)^{\frac{1}{2}}}$$

There is equality if the vectors $\vec{v}$ and $\Sigma_{YY}^{\frac{1}{2}} \Sigma_{YX} \Sigma_{XX}^{\frac{1}{2}} \vec{u}$ are colinear.

If $\vec{u}$ is the eigenvector with the maximum eigenvalue for the matrix $\Sigma_{XX}^{\frac{1}{2}} \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{YX} \Sigma_{XX}^{-\frac{1}{2}}$, the maximum of correlation is attained. Similarly, $\vec{v}$ is the eigenvector of $\Sigma_{YY}^{\frac{1}{2}} \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY} \Sigma_{YY}^{-\frac{1}{2}}$.

Hence, we have $\vec{w}_x$ is the eigenvector of $\Sigma_{XX}^{-1} \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{YX}$

$\vec{w}_y$ is the eigenvector of $\Sigma_{YY}^{-1} \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY}$.

(c) $\max Corr(\vec{w}_x^T X, \vec{w}_y^T Y) = \sqrt{\lambda_i} \quad \lambda_i$ is the ith eigenvalue of $\Sigma_{XX}^{-\frac{1}{2}} \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{YX} \Sigma_{XX}^{-\frac{1}{2}}$