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 4

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: Lagrange Multipliers

(a) $L(\theta,\lambda) = J(\theta) + \lambda\theta^T b$

$$= \sum_{k=1}^{n} \|\theta - x_k\|^2 + \lambda\theta^T b$$

$$\frac{dL(\theta,\lambda)}{d\theta} = 2\sum_{k=1}^{n}(\theta - x_k) + \lambda b$$

$$= 2n\theta - 2n\bar{x} + \lambda b$$

set $\frac{dL(\theta,\lambda)}{d\theta} = 0$

$$\theta = \frac{2n\bar{x} - \lambda b}{2n} = \bar{x} - \frac{\lambda b}{2n}$$

We have the constrain $\theta^T b = 0 \Longleftrightarrow b^T\theta = 0$

$$\frac{b^T 2n\bar{x} - \lambda b^T b}{2n} = 0$$

$$\lambda = \frac{2nb^T\bar{x}}{b^T b}$$

$$\theta = \bar{x} - \frac{\lambda b}{2n}$$

$$= \bar{x} - \frac{\frac{2nb^T\bar{x}}{b^T b} b}{2n}$$

$$= \bar{x} - \frac{b^T\bar{x}b}{b^T b}$$

(b) $L(\theta,\lambda) = J(\theta) + \lambda(\|\theta - c\|^2 - 1)$

$$= \sum_{k=1}^{n}\|\theta - x_k\|^2 + \lambda\|\theta - c\|^2 - \lambda$$

$$\frac{dL(\theta,\lambda)}{d\theta} = 2\sum_{k=1}^{n}(\theta - x_k) + 2\lambda(\theta - c)$$

$$= 2n\theta - 2n\bar{x} + 2\lambda\theta - 2\lambda c$$

$$= 2(n+\lambda)\theta - 2n\bar{x} - 2\lambda c$$

set $\frac{dL(\theta,\lambda)}{d\theta} = 0$

$$\theta = \frac{n\bar{x} + \lambda c}{n+\lambda}$$

We have the constrain $\|\theta - c\|^2 = 1$

$$\left\|\frac{n\bar{x}+\lambda c}{n+\lambda} - c\right\|^2 = 1$$

$$\left\|\frac{n\bar{x} - nc}{n+\lambda}\right\|^2 = 1$$

$$\left(\frac{n\bar{x} - nc}{n+\lambda}\right)^T\left(\frac{n\bar{x} - nc}{n+\lambda}\right) = 1$$

$$(n\bar{x}^T - nc^T)(n\bar{x} - nc^T) = (n+\lambda)^2$$

$$n^2\|\bar{x} - c\|^2 = (n+\lambda)^2$$

$$\lambda^2 + 2n\lambda + n^2 - n^2\|\bar{x} - c\|^2$$

$$\lambda = \frac{-2n \pm \sqrt{4n^2 - 4(n^2 - n^2\|\bar{x} - c\|^2)}}{2}$$

$$= -n \pm n\sqrt{\|x - c\|^2}$$

$$= -n \pm n\|x - c\|$$

$$\lambda_1 = -n - n\|\bar{x} - c\|$$

$$\lambda_2 = -n + n\|\bar{x} - c\|$$

$$\theta_1 = \frac{n\bar{x} + \lambda_1 c}{n + \lambda_1}$$

$$= \frac{n\bar{x} + (-n - n\|x - c\|)c}{n + (-n - n\|x - c\|)}$$

$$= \frac{n\bar{x} - nc - n\|x - c\|c}{-n\|x - c\|}$$

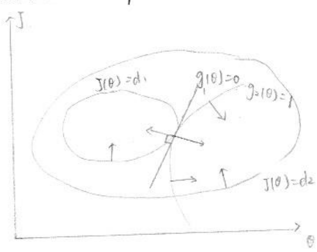
$$= \frac{\bar{x} - c - c\|x - c\|}{-\|x - c\|}$$

$$= \frac{c - \bar{x}}{\|x - c\|} + c$$

$$\theta_2 = \frac{n\bar{x} + \lambda_2 c}{n + \lambda_2}$$

$$= \frac{\bar{x} - c}{\|x - c\|} + c$$

Geometrical interpretation



(a) $g_1(\theta) = \theta^T b$ $\quad$ $\text{grad } J(\theta) = \lambda \text{ grad } g_1(\theta)$

(b) $g_2(\theta) = \|\theta - c\|^2$ $\quad$ $\text{grad } J(\theta) = \lambda \text{ grad } g_2(\theta)$

Exercise 2: Bounds on Eigenvalues

(a). ① from $x_1, \cdots, x_n \in R^d$. $m = \frac{1}{n}\sum_{k=1}^{n} x_k \in R^d$

We know S is real.

②. $\forall\ t = (t_1, \cdots, t_n)$

$$t S t^T = t \sum_{k=1}^{n} (x_k - m)(x_k - m)^T t^T$$
$$= \sum_{k=1}^{n} t (x_k - m)(x_k - m)^T t^T$$
$$= \sum_{k=1}^{n} (t(x_k - m))(t(x_k - m))^T$$
$$= \sum_{k=1}^{n} \| t(x_k - m) \|^2 \geq 0.$$

We can derive S is nonnegative

③ $|\lambda I - S| = 0$.

$$|\lambda I - S| = \lambda^d - \sum_{i=1}^{d} S_{ii} \lambda^{d-1} + \cdots = 0$$
$$\sum_{i=1}^{d} \lambda_i = -\frac{-\sum_{i=1}^{d} S_{ii}}{1} = \sum_{i=1}^{d} S_{ii}$$
$$\sum_{i=1}^{d} S_{ii} = \sum_{i=1}^{d} \lambda_i = \lambda_1 + \lambda_2 + \cdots + \lambda_d \geq \lambda_1$$

Thus $\sum_{i=1}^{d} S_{ii}$ is an upper bound to λ_1.

(b) The upper bound is $\sum_{i=1}^{d} \lambda_i$, if all other values $\lambda_2, \lambda_3, \cdots, \lambda_d$ are smaller, the bound is tighter. When $\lambda_2 = \lambda_3 = \cdots = \lambda_d = 0$, $\sum_{i=1}^{d} S_{ii} = \lambda_1$, the bound is tight. That means, only one feature is significant for the variance of the data, and all the data are in one line.

(c) We assume that $S_{jj} = \max_{i=1}^{d} S_{ii} > \lambda_1$, $j \in 1, 2, \cdots, d$.

$w^T S w = S_{jj} > \lambda_1 \geq \lambda_i$ where

$$w = (0, 0, \underset{\underset{j}{\uparrow}}{\cdots 1}, \cdots, 0)^T \quad \|w\| = 1$$

$$\lambda_i = \max_{w_i} w^T S w$$

Thus $\max_{i=1}^{d} S_{ii}$ is a lower bound to λ_1

(d) When $\lambda_1 = \max_{i=1}^{d} S_{ii}$, the lower bound is tight, which means the data is aligned one particular dimension.

Exercise 3: Iterative PCA

(a) $J(w) = \|Sw\| - \frac{1}{2} w^T S w$, $v = S^{0.5} w$

$$\frac{\partial J(w)}{\partial v} = \frac{\partial J}{\partial w} \cdot \frac{\partial w}{\partial v}$$

$$= \left(\frac{SSw}{\|Sw\|} - Sw \right) \cdot S^{-\frac{1}{2}}$$

$$= \frac{SS^{\frac{1}{2}} w}{\|Sw\|} - S^{\frac{1}{2}} w$$

$$= \frac{Sv}{\|Sw\|} - v$$

$$v \leftarrow v + \gamma \frac{\partial J}{\partial v}$$

$$= v + \gamma \left(\frac{Sv}{\|Sw\|} - v \right)$$

$$= v + \gamma \cdot \frac{Sv}{\|Sw\|} - \gamma v$$

if $\gamma = 1$

$$v \leftarrow v + \frac{Sv}{\|Sw\|}$$

(b) $\frac{\partial J}{\partial w} = \frac{SSw}{\|Sv\|} - Sw = 0$

Since S is invertible

$$\frac{Sw}{\|Sw\|} - w = 0$$

$$Sw = \|Sw\| w$$

$$\|Sw\| = \|Sw\| \cdot \|w\|$$

$$\|w\| = 1 .$$