

# Exercise Sheet 5

Machine Learning 2, SS16

May 27, 2016

Mario Tambos, 380599; Viktor Jeney, 348969; Sascha Huk, 321249; Jan Tinapp, 0380549

## Exercise 1 - Kernel Eigenvectors

(a)

We use Lagrange multipliers:

$$\begin{aligned}\mathcal{L} &= v^\top C v - \lambda(v^\top v - 1) \\ \frac{\partial \mathcal{L}}{\partial v} &= 2(C - \lambda I)v \stackrel{!}{=} \mathbf{0} \implies C v \stackrel{!}{=} \lambda v \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= 2(C - \lambda I)\end{aligned}$$

In every matrix  $C$  it holds that  $\text{eig}(C - \lambda I) = (\lambda_i - \lambda)_{i \in \{1, \dots, d\}}$  (see eq. 286 in matrix cookbook) where in this case  $\lambda = \max(\{\lambda_i\}_{i \in \{1, \dots, d\}})$ . This means that the second order derivative has only negative eigenvalues except for the largest eigenvalue, which is now zero. Hence  $\frac{\partial \mathcal{L}}{\partial v^2}$  is negative semi-definite (think of the bilinearform of combinations of eigenvectors, which correspond to the largest eigenvalue). We therefore use another simple criterion, namely using  $v^\top v = 1$  and the necessary condition  $C v \stackrel{!}{=} \lambda v$  in our objective function  $v^\top C v$ :

$$\max_v v^\top C v = \max(\{v^\top C v\}_v) = \max(\{v^\top \lambda_i v\}_{v,i}) = \max(\{\lambda_i\}_{i \in \{1, \dots, d\}})$$

Since the necessary condition wrt.  $v^\top v = 1$  describes an eigenvalue problem, both cases are equivalent.

(b)

$$\lambda v = C v \iff v = \frac{C v}{\lambda} \stackrel{\text{def. } C}{\iff} v = \sum_{i=1}^N \phi(x_i) \underbrace{\frac{\phi(x_i)^\top v}{\lambda}}_{\alpha_i}$$

Comparing the coefficients with the given equation  $v = \Phi^\top \alpha = \sum_{i=1}^N \phi(x_i) \alpha_i$ , we get:

$$\frac{\phi(x_i)^\top v}{\lambda} = \alpha_i \iff \phi(x_i)^\top v = \lambda \alpha_i \iff \Phi v = \lambda \alpha \quad (1)$$

We now conclude:

$$K \alpha = \Phi \Phi^\top \alpha = \Phi v \stackrel{(1)}{=} \lambda \alpha$$

## Exercise 2

Define

$$s := (s_i)_{i=1, \dots, N}$$

and the matrix of eigenvectors

$$U := (u_i^T)_{i=1, \dots, N}.$$

Then

$$Uy = s.$$

Because all eigenvectors are orthogonal,  $U$  is an orthogonal matrix. This means in particular, that

$$\|Uy\|_2^2 = \|y\|_2^2.$$

All in all we get

$$\sum_{i=1}^N s_i = \|s\|_2^2 = \|Uy\|_2^2 = \|y\|_2^2$$