# Exercise Sheet 4

Machine Learning 2, SS16

May 18, 2016

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

## Exercise 1 - Sparse Coding

**(a)**

$$\frac{\partial E}{\partial W} = \frac{\partial}{\partial W}\eta|W|_F^2 + \frac{\partial}{\partial W}\sum_{i=1}^{N}(|x^{(i)} - Ws^{(i)}|^2 + \lambda|s^{(i)}|_1)$$

$$= \eta\sum_{l}^{d}\sum_{k}^{h}\frac{\partial}{\partial W}(W_{lk})^2 + \sum_{i=1}^{N}\frac{\partial}{\partial W}(x^{(i)} - Ws^{(i)})^\top(x^{(i)} - Ws^{(i)})$$

$$= 2\eta W + \sum_{i=1}^{N} -2(x^{(i)} - Ws^{(i)}){s^{(i)}}^\top = 2\eta W - 2\sum_{i=1}^{N}(x^{(i)} - Ws^{(i)}){s^{(i)}}^\top$$

**(b)**

$$\frac{\partial E}{\partial s^{(i)}} = \frac{\partial}{\partial s^{(i)}}\eta|W|_F^2 + \frac{\partial}{\partial s^{(i)}}\sum_{j=1}^{N}(|x^{(j)} - Ws^{(j)}|^2 + \lambda|s^{(j)}|_1)$$

$$= \frac{\partial}{\partial s^{(i)}}(x^{(i)} - Ws^{(i)})^\top(x^{(i)} - Ws^{(i)}) + \frac{\partial}{\partial s^{(i)}}\lambda|s^{(i)}|_1$$

$$= -2W^\top(x^{(i)} - Ws^{(i)}) + \lambda\sum_{k=1}^{h}\frac{\partial}{\partial s^{(i)}}{s^{(i)}}_k \qquad ({s^{(i)}}_k \geq 0)$$

$$= -2W^\top(x^{(i)} - Ws^{(i)}) + \lambda 1_h$$

## Exercise 2 - Sparsifying Non-Linearities

**(a)** The derivative wrt. to W is already equivalent. Taking the derivative wrt. $r^{(i)}$ we obtain:

$$\frac{\partial}{\partial r^{(i)}}\sum_{j=1}^{N}(|x^{(j)} - Wg(r^{(j)})|^2 + \lambda|r^{(j)}|^2) = \frac{\partial}{\partial r^{(i)}}|x^{(i)} - Wg(r^{(j)})|^2 + \frac{\partial}{\partial r^{(i)}}\lambda|r^{(i)}|^2$$

$$= -2W^\top(x^{(i)} - Wg(r^{(i)}))\frac{\partial}{\partial r^{(i)}}g(r^{(i)}) + \lambda 2r^{(i)}$$

Comparing the factors of $\lambda$ yields componentwise differences by the componentwise factors $2{r^{(i)}}_k$ $\forall k \in \{1, \ldots, h\}$. In order for the two problems to be equal, we choose g such that it's Jacobian is diagonal and such that it has the k-th of these factors on the k-th diagonal element. Therefore we have an equivalent problem for $g(r^{(i)}) = ({r^{(i)}}_1^2, \ldots, {r^{(i)}}_h^2)$.