Machine Learning	Dr. Mario Fritz and Dr. Björn Andres
Exercise Sheet 10 - 19.01.2015	due: 26.01.2015

Exercise 23 - Semi-supervised Learning

a. **(4 Points)** Prove the following equality:

$$\left\langle f, (\mathbb{1} - D^{-\frac{1}{2}}\, W\, D^{-\frac{1}{2}}) f \right\rangle_{\mathbb{R}^n} = \frac{1}{2} \sum_{i,j=1}^{n} w_{ij} \Big(\frac{f_i}{\sqrt{d_i}} - \frac{f_j}{\sqrt{d_j}} \Big)^2 .$$

b. **(4 Points)** In the lecture it was shown that the solution of

$$\underset{f \in \mathbb{R}^n}{\arg\min} \quad \sum_{i \in T} (y_i - f_i)^2 + \frac{\lambda}{2} \sum_{i,j \in T} w_{ij} \left(\frac{f_i}{\sqrt{d_i}} - \frac{f_j}{\sqrt{d_j}} \right)^2 ,$$

can be written as

$$f^* = (1 - \alpha) \sum_{r=0}^{\infty} \alpha^r S^r Y.$$

Derive the solution f^* in the limit of $\lambda \to 0$ (resp. $\alpha \to 0$) where the graph is sparse (in particular it is not fully connected), that is derive for each vertex i its label (note that the final classification of vertex i is done with $\mathrm{sign}(f^*(i))$). For simplicity you are allowed to replace S with the weight matrix W.

Hints:

a. For the second part you have to consider the limit $\lambda \to 0$. Setting $\lambda = 0$ will yield a different result !
It is helpful to expand the sum in the solution $f^*(i)$ for one vertex i. What happens as $\alpha \to 0$? Which terms in the sum are non-zero ? Which terms dominate the solution as $\alpha \to 0$? What is the meaning of $(W^r)_{ij}$ (think of paths in the graph) ?

Exercise 24 - Optimal Approximation

Given a set of points $\{X_i\}_{i=1}^n$ in $\mathbb{R}^d$.

a. **(2 Points)** Which is the point which approximates best the set of n points by minimizing the sum of squared Euclidean distances. In other words, solve the following problem

$$Z^* = \underset{Z \in \mathbb{R}^d}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \|X_i - Z\|^2 .$$

b. **(2 Points)** Given that you have a unit vector $v \in \mathbb{R}^d$. What is the meaning of the linear mapping, $A : \mathbb{R}^d \to \mathbb{R}^d$, defined as

$$Au = v\, v^T\, u.$$

Show that $A^2 = A$ (idempotent) and $A^T = A$ (symmetric).

c. **(2 Points)** Which linear subspace is the best approximation to the set of n points in the sense that the squared Euclidean distance to the subspace is minimized.

$$v^* = \underset{v \in \mathbb{R}^d,\, \|v\|=1}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \left\| X_i - (vv^T) X_i \right\|^2 .$$