ICS-E4030 Kernel methods in machine learning, course exam 14.12.2015 / Examiner: Juho Rousu

Instructions:

- *You have 3 hours to complete exam.*
- *You are allowed to use one two-sided cheat-sheet (A4 page, both sides handwritten), which you have to submit together with the exam paper.*
- *You may use a scientific calculator.*
- *No additional material is allowed.*

Questions

1. (10 points) Give short (a few sentences) definitions of the following concepts.

   (a) Ridge regression
   (b) Sub-gradient
   (c) Kendall's distance
   (d) Semantic kernel
   (e) Pre-image problem

2. (10 points) Prove the following proposition:

   Consider an undirected graph $G = (V, E, W)$ with vertices $V = \{v_1, \ldots, v_n\}$, edges $E \subset V \times V$ and edge weights $W = (w_{ij})_{(i,j) \in E}, w_{ij} \geq 0$.
   The unnormalized graph Laplacian matrix on this graph is defined as

   $$L = D - W.$$

   where D is a diagonal matrix with diagonal elements $D_{ii} = \sum_j w_{ij}$,
   Prove the following properties:

   a) (2.5 points) For every vector $f \in \mathcal{R}^n$, we have $f'Lf = \frac{1}{2} \sum_{i,j=1}^{n} w_{ij}(f_i - f_j)^2$.
   b) (2.5 points) L is symmetric and positive semi-definite.
   c) (2.5 points) If the graph is connected, then there is only one eigenvalue equal to 0 with all ones vector (up to a constant) as the eigenvector.
   d) (2.5 points) If there are k connected components in the graph, then we have k eigenvalues equal to 0 with corresponding eigenvectors as indicator vectors (up to a constant) of those components.

3. (10 points) Explain in detail what are graph kernels and how they can be used in machine learning. Give examples of at least two graph kernels.

4. (10 points) Derive the dual problem of the structured output optimization problem (soft-margin case), where $\mathcal{Y}$ is a finite space of outputs, $\varphi(x_i, \mathbf{y}_i)$ is the joint feature map, $\ell(\mathbf{y}_i, \mathbf{y})$ is the loss function, $\mathbf{w}$ is the weight vector, and ξ_i is the slack allocated to training example x_i:

$$\begin{aligned} \min_{\mathbf{w}, \xi_i \geq 0} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{m} \xi_i \\ s.t. \quad & \langle \mathbf{w}, \varphi(x_i, \mathbf{y}_i)\rangle - \langle \mathbf{w}, \varphi(x_i, \mathbf{y})\rangle \geq \ell(\mathbf{y}_i, \mathbf{y}) - \xi_i,\ i = 1, \ldots, N, \mathbf{y} \in \mathcal{Y} \end{aligned}$$

5. (10 points) Explain in detail how multiple input kernels can be combined and used in classification tasks.