Exercises
**Data Mining: Learning from Large Data Sets**
FS 2016

**LAS Group, Institute for Machine Learning**
Dept. of Computer Science, ETH Zürich
**Prof. Dr. Andreas Krause**
Web: `http://las.ethz.ch/courses/dm-f16/`
**Email questions to:**
Kfir Levy, yehuda.levy@inf.ethz.ch

# Series 4, Nov 10th, 2016 (SVM / Kernel)

**It is not mandatory to submit solutions and sample solutions will be published after one week. If you choose to submit your solution, please send an e-mail from your `ethz.ch` address with subject `Exercise4` containing a PDF (LaTeXor scan) to `yehuda.levy@inf.ethz.ch` until Wednesday, Nov 16th 2016.**

### Problem 1 (Support Vector Machines):

The objective of this exercise is to investigate the $L2$-SVM which uses the square sum of the slack variables $\xi_i$ in the objective function instead of the linear sum of the slack variables (i.e. squaring the hinge loss). Let $S = \{(x_1, y_1), \cdots, (x_n, y_n)\}$ be a training set of examples and binary labels $y_i \in \{-1, +1\}$. The primal formulation of the $L2$-SVM is as follows

$$\min_{\mathbf{w},\xi} \tfrac{1}{2}\|\mathbf{w}\|^2 + \tfrac{C}{2}\sum_{i=1}^{n} {\xi_i}^2$$

s.t.

$$\begin{array}{ll} y_i \mathbf{w}^T \mathbf{x}_i \geq 1 - \xi_i & i = 1, \ldots, n \\ \xi_i \geq 0 & i = 1, \ldots, n \end{array} \tag{1}$$

- Reformulate the above optimization as an unconstrained optimization problem.
- Give a step-by-step solution to deriving the optimal parameters using stochastic gradient descent.

### Problem 2 (Deriving the SVM Dual):

Consider the following SVM formulation:

$$\text{minimize}_{w,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i \tag{2}$$

subject to

$$y_i \mathbf{w}^T \mathbf{x} \geq 1 - \xi_i \text{ for all } i = 1, \cdots, n \tag{3}$$

and

$$\xi_i \geq 0 \text{ for all } i = 1, \cdots, n \tag{4}$$

- Write down the Lagrangian using $\alpha_i$ as the Lagrange multiplier corresponding to constraint 3 and $\gamma_i$ as the Lagrange multiplier corresponding to constraint 4.

- Compute the derivative of the Lagrangian with respect to **w** and $\xi_i$.
- Solve for the dual and show it is given by

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j. \tag{5}$$

subject to

$$0 \leq \alpha_i \leq C \text{ for } i = 1, \cdots, n \tag{6}$$

**Problem 3 (Kernelized Ridge Regression):**

Consider the following ridge regression problem:

$$\min_{\mathbf{w}} \sum_{i=1}^{n} ||y_i - \mathbf{w}^T \mathbf{x}_i||_2^2 + \lambda ||\mathbf{w}||_2^2,$$

- Compute the derivative of the above objective function with respect to **w**, and derive its closed form solution.
- Show that the closed form solution you derived in the previous step can be written as

$$\mathbf{w} = \mathbf{X}^T (\mathbf{X}\mathbf{X}^T + \lambda \mathbf{I})^{-1} \mathbf{y}.$$

- Substitute $\mathbf{w} = \mathbf{X}^T \alpha$ to kernelize the above ridge regression problem.