ENGN 4520: Introduction to Machine Learning

Alex Smola, RSISE ANU
Problem Sheet — Week 5
The due date for these problems is Thursday, June 7

Teaching Period
April 30 to June 8, 2001

A Theory

Problem 18 (Scale Invariance for Classification, 6 Points)
Show that for classification using the ν-trick, the following two optimization problems are equivalent:

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + \sum_{i=1}^{m}(\xi_i - \nu\rho) \\ \text{subject to} \quad & y_i(\langle \mathbf{w}, \mathbf{x}_i\rangle + b) \geq \rho - \xi_i \text{ and } \xi_i \geq 0 \end{aligned}$$

and

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{m}(\xi_i - \nu\rho) \\ \text{subject to} \quad & y_i(\langle \mathbf{w}, \mathbf{x}_i\rangle + b) \geq \rho - \xi_i \text{ and } \xi_i \geq 0 \end{aligned}$$

In other words, show that one can transform the solution $(\mathbf{w}, b)$ of the first problem into one of the second $(\mathbf{w}', b')$ by linear scaling.

Problem 19 (Stochastic Gradient Descent with ν, 7 Points)
Denote by $R_{\text{reg}}[f]$ the regularized risk functional for a linear model $f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x}\rangle + b$ with

$$R_{\text{reg}}[f] = \frac{1}{m}\sum_{i=1}^{m}\max(0, \rho - y_i f(\mathbf{x}_i)) + \varepsilon\nu + \frac{\lambda}{2}\|\mathbf{w}\|^2$$

in a classification setting ($y_i \in \{-1, 1\}$).

1. *Derive a stochastic gradient descent algorithm in $\mathbf{w}, b$.*
2. *Rewrite the algorithm using feature space methods, i.e. $\mathbf{x} \to \Phi(\mathbf{x})$ and kernels.*
3. *Give a kernel expansion for f.*

Problem 20 (Adaptive Margin Width in Regression, 7 Points)
Assume that we want to perform regression on $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)$ with the ε-insensitive loss function

$$c(\mathbf{x}, y, f(\mathbf{x})) = |y - f(\mathbf{x})|_\varepsilon$$

1. *Set up the equations for the regularized risk functional using c as a loss function.*
2. *Modify the regularized risk functional such that for each observation $(\mathbf{x}_i, y_i)$ we may use a different width ε_i.*
3. *Rewrite the regularized risk functional as a constrained quadratic optimization problem.*
4. *Compute the dual optimization problem.*

Alex Smola, RSISE ANU
Problem Sheet — Week 5
Teaching Period
April 30 to June 8, 2001

The due date for these problems is Thursday, June 7

B Programming

Problem 21 (Ugly Digits, 20 Points)
The goal is to use the novelty detection algorithm with ν, as describes in the lecture, to find badly written digits in a database with handwritten digits. You need a computer with **64MB memory** *and a new version of SVLab for this problem.*

1. *Download the dataset from* `http://axiom.anu.edu.au/~smola/engn4520/usps.zip`*, unzip it and load it into MATLAB. It contains a training set (2000 examples) and a test set (6000 examples).*
2. *Describe (on paper) the variable assignment for a quadratic optimizer of your choice.*
3. *Implement in MATLAB an algorithm that takes $(\mathbf{x}_1, \ldots, \mathbf{x}_m)$, ν, and the kernel as an input and outputs α_i and ρ, which solves the novelty detection problem.*
4. *Rank your training set according the function values $f(\mathbf{x}_i)$, i.e. worst examples first for a Gaussian RBF kernel*
$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x} - \mathbf{x}'\|^2\right)$$
with the following values for σ^2: $\sigma^2 = 0.2d, \sigma^2 = 0.5d, \sigma^2 = d$ where $d = 256$ and $\nu = 0.01$ and $\nu = 0.1$ for the current dataset.
5. *Rank your test set according to the same criterion. Plot the images of the so-chosen samples.* **Hint: you can reshape the vector $\mathbf{x}_i$ into a 16×16 matrix with the `reshape` command in MATLAB.**

More Instructions for SVLab Toolbox

Download the new version of SVLab `http://axiom.anu.edu.au/~smola/engn4520/svlab.zip` and install it on your system. In addition to the functions of previous week you will need to use the `intpoint_pr` object which is a quadratic optimizer (you can also use the built-in `qp` routine of MATLAB if your computer has the license). It is used as follows

```
optimizer = intpoint_pr; %creates optimizer object (intpoint_re is for regression)
optimizer.verbose = 2;   %makes the output verbose (recommended)

[primal, dual, how] = optimize(optimizer, c, H, A, b, l, u);
%calls the optimizer with quadratic matrix $H$, linear term $c$, constraint
%matrix $A$, lower bound $l$, and uppper bound $u$
```

This solves the following optimization problem:

$$\begin{aligned} \text{minimize} \quad & c^\top x + \frac{1}{2} x^\top H x \\ \text{subject to} \quad & l_i \leq x_i \leq u_i \text{ and } Ax = b \end{aligned}$$