THE AUSTRALIAN NATIONAL UNIVERSITY
First Semester Examinations 2001

ENGN4520 ENGINEERING S1
INTRODUCTION TO MACHINE LEARNING

Writing Period: 2 hours duration

Study Period: 15 minutes duration

Permitted Materials: calculator (non-programmable), course/lecture notes, dictionary, mathematical tables, notes, slide rules, statistical tables

Note: You do not have to solve all problems to obtain full points. There are more problems than you can solve within 2 hours.

Problem 1 (Cancer Diagnosis, 10 Points)

Assume a patient visits a physician to undergo a test whether he has a certain type of cancer or not. The probability that an arbitrary person will develop this type of cancer is 1 in 100,000.

1. *The patient tests positive (= cancer) to a test with the following properties: in 99 out of 100 cases it detects cancer if it is present. However, it raises a false alarm in 2 out of 1000 cases.*

   What is the probability of cancer, given that the test is positive.

2. *Additionally, the patient exhibits a certain symptom that is typically associated with cancer. In particular, all cancer patients show this symptom, however in one of 500 cases another disease is responsible for this symptom. Assume that the presence of this symptom is* independent *of the outcome of the previous test.*

   What is the probability of cancer now, given the knowledge about the symptom and *the test.*

Problem 2 (Weighting Patterns with Support Vectors, 15 Points)
Assume we have a linear Support Vector Machine with soft margin loss, i.e.

$$c(\mathbf{x}, y, f(\mathbf{x})) = \max(0, 1 - yf(\mathbf{x})) \tag{1}$$

which is to be trained on some training data $(\mathbf{x}_1, y_1), \ldots (\mathbf{x}_m, y_m)$. *Moreover assume that we know that some of the observations* $(\mathbf{x}_i, y_i)$ *are more important than others, specifically that there exist weighting coefficients* $C_i > 0$ *such that we minimize a modified regularized risk functional*

$$\sum_{i=1}^{m} C_i c(\mathbf{x}_i, y_i, f(\mathbf{x}_i)) + \frac{1}{2}\|w\|^2. \tag{2}$$

1. *Rewrite eq. (2) such that it becomes a constrained quadratic optimization problem, i.e. with linear constraints and a quadratic objective function.*
2. *Derive the Lagrange function corresponding to the constrained optimization problem.*
3. *Compute the dual optimization problem.*
4. *Compare the result to the standard soft margin Support Vector Machine.*

Problem 3 (Perceptron Algorithm and Stochastic Gradient Descent, 15 Points)

We use a loss function $c(\mathbf{x}, y, f(\mathbf{x})) = |y - f(\mathbf{x})|_\varepsilon$ *and want to minimize the* **empirical risk** $R_{\text{emp}}[f]$.

1. *Compute derivatives of* c *with respect to* $\mathbf{w}$, b *for a linear model* $f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b$.
2. *State the general gradient descent rule regarding* $\mathbf{w}, b$ *for arbitrary* c.
3. *State the gradient descent rule for c specified as above.*
4. *With a learning rate* $\eta = 0.1$ *and* $\varepsilon = 0.5$ *calculate the values of* $\mathbf{w}, b$ *for the first four steps. The data set of* (x, y) *is given by* $(1, 1), (3, 5), (2, 3), (1, 2)$ *and we initialize* $\mathbf{w} = \mathbf{0}$ *and* $b = 0$.
5. *Modify the stochastic gradient algorithm using* $|y - f(\mathbf{x})|_\varepsilon$ *as a loss function such that it works in feature space using kernels, i.e. using* $\Phi(\mathbf{x})$ *and* $k(\mathbf{x}, \mathbf{x}')$.

Problem 4 (Admissible Kernel, 15 Points)

1. *Show that the matrix* $K_{ij} := y_i y_j k(\mathbf{x}_i, \mathbf{x}_j)$ *is positive semidefinite if* k *satisfies Mercer's condition.*

2. *Show that if* k_1, k_2 *satisfy Mercer's condition, then also any* $k = \alpha_1 k_1 + \alpha_2 k_2$ *with* $\alpha_1, \alpha_2 \geq 0$ *also satisfy Mercer's condition.*

3. *Show that the kernel* $k(\mathbf{x}, \mathbf{x}') = \kappa(\langle \mathbf{x}, \mathbf{x}' \rangle)$ *with* $\kappa(\xi) = \sum_{l=1}^{n} c_l \xi^l$ *and* $c_l \geq 0$ *satisfies Mercer's condition.*

4. *Show that the kernel* $k(x, x') = \frac{1}{1+(x-x')^2}$ *with* $x, x' \in \mathbb{R}$ *satisfies Mercer's condition. Discuss its regularization properties.*

Problem 5 (Hilbert Spaces, 15 Points)
We begin with the quadratic form on functions $f : \mathbb{R} \to \mathbb{R}$.

$$q(f) := \|f\|^2 + 2\|f'\|^2 + 3\|f''\|^2 \tag{3}$$

1. *Use the polarization inequality to recover the dot product underlying the definition of the quadratic form* $q(f)$*, i.e. find a bilinear form from* $q(f)$ *such that* $\langle f, f \rangle_{\mathcal{H}} = q(f)$*. Show that what you obtained is a dot product.*

2. *Compute* $q(f)$ *for* $f = c_0 + \displaystyle\sum_{l=1}^{n} (s_l \sin(lx) + c_l \cos(lx))$.

3. *Compute the representation of* $q(f)$ *in the Fourier domain, i.e. in terms of* $\tilde{f}(\omega)$.

4. *Show that the corresponding kernel in feature space setting is translation invariant, i.e.* $k(x, x') = k(x - x')$. **Note: you need not compute** k **for that purpose!**