UNIVERSITY COLLEGE LONDON

EXAMINATION FOR INTERNAL STUDENTS

MODULE CODE	:	**COMPGI13**
ASSESSMENT PATTERN	:	**COMPGI13B**
MODULE NAME	:	**Advanced Topics in Machine Learning**
DATE	:	**23-May-13**
TIME	:	**10:00**
TIME ALLOWED	:	**2 Hours 30 Minutes**

Advanced Topics in Machine Learning, GI13, 2012/13

Answer any THREE questions. Each question is worth 20 marks. Use separate answer books for PART A and PART B. **Gatsby PhD students only**: answer *either* TWO questions from PART A and ONE question from PART B; *or* ONE question from PART A and TWO questions from PART B.

Marks for each part of each question are indicated in square brackets

Calculators are NOT permitted

Part A: Kernel Methods

1. We define $\mathcal{F}$ to be a reproducing kernel Hilbert space on domain $\mathcal{X}$ with feature map $\phi(x)$ and kernel $k(x,x') := \langle \phi(x), \phi(x') \rangle_{\mathcal{F}}$, and $\mathcal{G}$ to be a reproducing kernel Hilbert space on domain $\mathcal{Y}$ with feature map $\psi(y)$ and kernel $l(y,y') := \langle \psi(y), \psi(y') \rangle_{\mathcal{G}}$.

   The Hilbert-Schmidt operators mapping from $\mathcal{F}$ to $\mathcal{G}$ form a Hilbert space, written $\mathrm{HS}(\mathcal{F}, \mathcal{G})$, with inner product

$$\langle L, M \rangle_{\mathrm{HS}} = \sum_{i \in I} \langle Le_i, Me_i \rangle_{\mathcal{G}}, \tag{1}$$

   where $\{e_i\}_{i \in I}$ form an orthonormal basis for $\mathcal{F}$. It follows that the norm for elements in this Hilbert space is written

$$\|L\|^2_{\mathrm{HS}} = \sum_{i \in I} \|Le_i\|^2_{\mathcal{G}} \tag{2}$$

   (a) Given $a \in \mathcal{F}$ and $b \in \mathcal{G}$, we define the tensor product $b \otimes a$ as a rank-one operator from $\mathcal{F}$ to $\mathcal{G}$,

$$(b \otimes a) f \mapsto \langle f, a \rangle_{\mathcal{F}} b.$$

   Show that

$$\|b \otimes a\|_{\mathrm{HS}} = \sqrt{\|a\|^2_{\mathcal{F}} \|b\|^2_{\mathcal{G}}}.$$

   You will need Parseval's identity:

$$\sum_{i \in I} \left| \langle a, e_i \rangle_{\mathcal{F}} \right|^2 = \|a\|^2_{\mathcal{F}}.$$

[4 marks]

(b) Show that

$$\langle L, b \otimes a \rangle_{\text{HS}} = \langle b, La \rangle_{\mathcal{G}} \tag{3}$$

for all $L \in \text{HS}(\mathcal{F}, \mathcal{G})$. Hint: begin by expanding a in terms of the orthonormal basis $\{e_i\}_{i \in I}$.

[5 marks]

(c) Consider the operator from $\text{HS}(\mathcal{F}, \mathcal{G})$ to $\mathbb{R}$, defined

$$T_{YX}(A) := E_{XY} \langle \psi(Y) \otimes \phi(X), A \rangle_{\text{HS}}.$$

The operator T_{YX} is bounded as long as

$$|T_{YX}(A)| \leq \lambda_{YX} \|A\|_{HS}$$

for $\lambda_{YX} < \infty$. Under what condition is the operator T_{YX} bounded? Express this in terms of expectations of kernel functions.You will need a result following from Jensen's inequality,

$$|E_X x| \leq E_X |x|,$$

the Cauchy-Schwarz inequality,

$$|\langle a, b \rangle| \leq \|a\| \|b\|,$$

and the result from the first part.

[3 marks]

(d) The Riesz representer theorem states there exists an element $C_{YX} \in \text{HS}(\mathcal{F}, \mathcal{G})$ such that the bounded linear operator $T_{YX} : \text{HS}(\mathcal{F}, \mathcal{G}) \to \mathbb{R}$ can be written

$$T_{YX}(A) = \langle C_{YX}, A \rangle_{\text{HS}}. \tag{4}$$

Show the uncentred covariance operator C_{YX} has the property

$$\langle g, C_{YX} f \rangle_{\mathcal{G}} = E_{XY} [f(X) g(Y)].$$

Hint: use the result from the second part.

[4 marks]

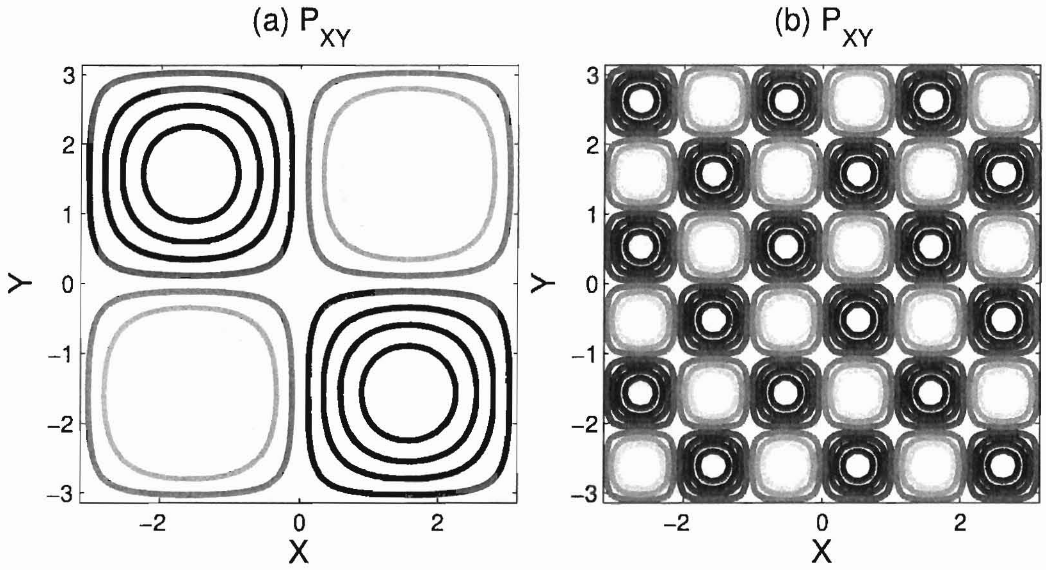

Figure 1: Contours of two probability density functions.

(e) The constrained covariance is the maximum singular value of C_{YX}, i.e.

$$\mathrm{COCO} = \max_{\|f\|_{\mathcal{F}\leq 1}, \|g\|_{\mathcal{G}\leq 1}} \langle g, C_{YX} f \rangle_{\mathcal{G}}.$$

In figure (1), two density functions are shown as contour plots. Assume both k and l are Gaussian kernels. Which of the two density functions (a) and (b) will have the largest COCO? Provide a short qualitative explanation as to why.

[4 marks]

2. Independent component analysis involves the recovery of sources which have been linearly mixed, based only on their mutual independence:

$$\mathbf{x} = M\mathbf{s},$$

where $\mathbf{x}$ is a $d \times 1$ vector, M is a $d \times d$ matrix, and $\mathbf{s}$ is a $d \times 1$ vector with independent components, having probability density function

$$P_S(\mathbf{s}) = \prod_{i=1}^{d} P_{S_i}(s_i).$$

Here the subscript S in $P_S(\mathbf{s})$ is used to identify the probability density, and the argument $\mathbf{s}$ is the value taken by the random variable. The recovered sources are written

$$\mathbf{y} = B\mathbf{x},$$

hence B is an estimate of M^{-1} (up to certain indeterminacies).

(a) Give two examples of linear transformations which cannot be recovered using independence.

[3 marks]

(b) Assume we make n independent, identically distributed observations $\{\mathbf{x}_i\}_{i=1}^{n}$, giving the matrix

$$X = \begin{bmatrix} \mathbf{x}_1 & \dots & \mathbf{x}_n \end{bmatrix}$$

We define a linear transformation B_w such that $T = B_w X$. Assuming the X have zero mean, find the matrix B_w such that T is decorrelated,

$$C_t = TT^\top = I,$$

where C_t is the empirical covariance of T. What linear operation remains to be done to recover the sources, following this decorrelation step?

[3 marks]

(c) You are doing ICA, where you want to recover a 2×1 vector of independent sources $\mathbf{s}$ from mixtures $\mathbf{x}$, assuming that

$$\mathbf{x} = M\mathbf{s},$$

where M is an orthogonal matrix, $M^\top M = I$. Explain why the mixing matrix M is not recoverable if the sources $\mathbf{s}$ are independent Gaussian random variables with zero mean and unit variance. A proof is needed for full credit on this question.

[4 marks]

(d) We are given an assumed source probability density model $\widehat{P}_S = \prod_{i=1}^{d} \widehat{P}_{S_i}$. Assume the mixing matrix M is *orthogonal*. State the expected log likelihood of the observations X, as a function of the estimated inverse B_r of M and of the assumed source densities $\widehat{P}_{S_i}$. You will need the results:

$$\begin{aligned} \mathbf{x} &= A\mathbf{s} \\ P_X(\mathbf{x}) &= \det(A^{-1}) P_S(A^{-1}\mathbf{x}) \end{aligned} \tag{5}$$

for some matrix A, and $\det(Q) = 1$ for orthogonal Q.

[2 marks]

(e) Another option for recovering independent sources is to minimize the mutual information. Again assume the mixing matrix M is orthogonal. Writing as

$$\mathbf{y} = B_r \mathbf{x} \tag{6}$$

the recovered sources, where the estimated unmixing matrix B_r is orthogonal, the mutual information is

$$I(\mathbf{y}) = \int \log\left(\frac{P_Y(\mathbf{y})}{\prod_{\ell=1}^{d} P_{Y_\ell}(y_\ell)}\right) P_Y(\mathbf{y}) d\mathbf{y}$$

Using equations (5) and (6), show that minimizing the mutual information is the same as the maximum likelihood when M and B_r are orthogonal, as long as the model $\widehat{P}_S$ is correct.

[5 marks]

(f) A lecturer demonstrates ICA by recovering a linear mixture of the two signals in Figure 2. Why is this an incorrect use of ICA?

[3 marks]

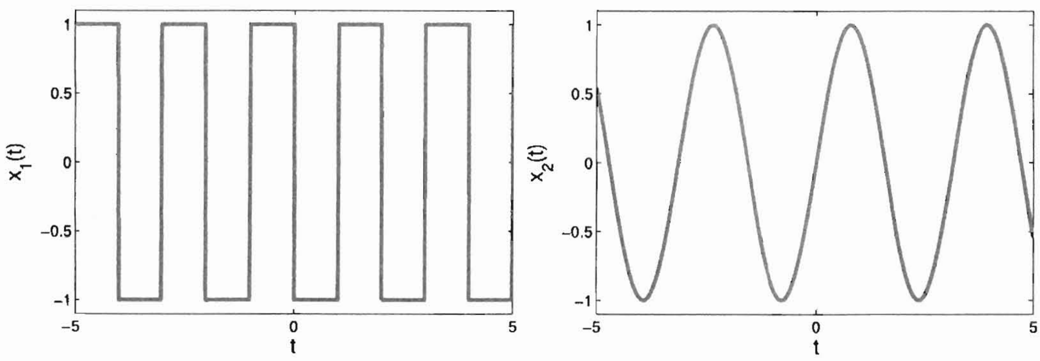

Figure 2: Two signals x_1 and x_2 which are proposed to be recovered via ICA, following linear mixing.

3. Assume $\mathcal{H}$ is a reproducing kernel Hilbert space with a Gaussian kernel,

$$k(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{\sigma}\right) = \langle \phi(x_i), \phi(x_j) \rangle_{\mathcal{H}} \tag{7}$$

We have a sample $(x_i, y_i)_{i=1}^n$ drawn independently and identically from some distribution P_{XY}, where the $y_i \in \mathbb{R}$. Support vector regression finds a function:

$$f(x) = \langle w(\cdot), \phi(x) \rangle_{\mathcal{H}}$$

which solves the following problem:

$$\underset{w \in \mathcal{H}, \xi \in \mathbb{R}^n}{\text{minimize}} \frac{1}{2}\|w\|_{\mathcal{H}}^2 + \frac{C}{m}\sum_{i=1}^{m}(\xi_i + \xi_i^*), \tag{8}$$

$$\text{subject to} \quad (\langle w(\cdot), \phi(x_i) \rangle_{\mathcal{H}} + b) - y_i \leq \varepsilon + \xi_i \tag{9}$$

$$y_i - (\langle w(\cdot), \phi(x_i) \rangle_{\mathcal{H}} + b) \leq \varepsilon + \xi_i^* \tag{10}$$

$$\xi_i, \xi_i^* \geq 0$$

where $C, \varepsilon \in \mathbb{R}^{++}$ are parameters of the algorithm (the notation means that both C and ε are strictly greater than zero).

(a) Define strong duality in the general setting of an optimization problem with equality and inequality constraints. Then describe the two conditions that hold for the support vector regression problem which ensure strong duality (hint: the second of these conditions is trivially satisfied here, since there are no equality constraints).

[2 marks]

(b) Write the Lagrangian for the SV regression problem. State the KKT conditions as they apply to the problem. What is implied about the maximum of the dual problem when the KKT conditions hold?

[6 marks]

(c) Write the Lagrange dual function for this optimization problem. In particular, you should obtain a form

$$w = \sum_{i=1}^{m} (\alpha_i^* - \alpha_i)\phi(x_i)$$

due to the two constraints (9) and (10).

[6 marks]

(d) What do the KKT conditions imply about the allowable range of α_i? Where are points with $\alpha_i = 0$ situated relative to the regression function $f(x)$? Where are points for which α_i attains its upper bound? Finally, where are those points with α_i between the lower and upper bound (you do *not* need to obtain the analogous results for α_i^*)?

[6 marks]

Part B: Reinforcement Learning

4. A squirrel is searching for food in three trees: an Aspen, a Birch and a Cherry tree. The squirrel's energy is measured in *squizzles*. If it is in either the Aspen or Birch tree, it can *jump* down to the Cherry tree, at a cost 2 of squizzles. If it is in the Cherry tree, it can *gather* cherries, with an expected gain of 2 squizzles; or it can *climb* up to the Aspen tree, which costs 4 squizzles. If it is in the Aspen tree, it can also *balance* along a branch, to collect 4 squizzles of nuts, which then leads either to the Birch tree with probability 1/2 or to the Cherry tree with probability 1/2.

   (a) Formalise the above description as a Markov Decision Process (MDP) with discount factor $\gamma = 0.5$. Draw a diagram showing the states, transitions, and rewards for the squirrel MDP.

   [5 marks]

   (b) Write down the *Bellman expectation equation* for state-value functions.

   [2 marks]

   (c) The squirrel starts with a uniform random policy π_1 that takes all available actions with equal probability. Write down three equations corresponding to the Bellman equations for $V^{\pi_1}(Aspen)$, $V^{\pi_1}(Birch)$, and $V^{\pi_1}(Cherry)$.

   [3 marks]

   (d) Solve your equations to find the state-value function $V^{\pi_1}(Aspen)$, $V^{\pi_1}(Birch)$, and $V^{\pi_1}(Cherry)$.

   [3 marks]

   (e) Apply one iteration of greedy policy improvement to compute a new, deterministic policy π_2. Using your new policy, which action will the squirrel select from the Aspen tree and the Cherry tree?

   [2 marks]

   (f) Write down the *Bellman optimality equation* for state-value functions.

   [2 marks]

   (g) Starting with an initial value function of $V_1(Aspen) = V_1(Birch) = V_1(Cherry) = 1$, apply one iteration of value iteration (i.e. one backup for each state, synchronously) to compute a new value function $V_2(s)$.

[3 marks]

[Total 20 marks]

5. A mouse is in an experiment involving sound. At each time-step a buzzer is either ON or OFF. At the end of each episode the mouse receives a reward. Consider the following three episodes:

Episode	$t=1$	$t=2$	$t=3$	reward
A:	OFF	OFF	ON	$+2$
B:	ON	ON	OFF	$+2$
C:	OFF	ON	ON	$+1$

The mouse would like to predict the reward from the buzzer sequence up to and including time-step t. Assume an undiscounted reward process, i.e. there are no actions and the discount factor is $\gamma = 1$.

Consider a linear function approximator $V(h_t) = N(h_t)\theta$, where $N(h_t)$ is the total number of times that the buzzer has been ON in the sequence h_t, up to and including time-step t, and θ is a scalar parameter.

(a) What are the values of $N(h_t)$ at each step t of these three episodes?

[2 marks]

(b) What are the returns v_t at each step t of these three episodes?

[2 marks]

(c) The mean-squared error is $\varepsilon^2(\theta) = \mathbb{E}_\pi\left[(V(h_t) - v^t)^2 \mid s = s_t\right]$. What is the gradient of ε^2 w.r.t. θ?

[2 marks]

(d) Use your gradient to derive a Monte-Carlo policy evaluation algorithm. Specifically, give an update equation that incrementally modifies θ by stochastic gradient descent, after observing a return v_t following from a history h_t.

[2 marks]

(e) Use your update equation for incremental Monte-Carlo policy evaluation to derive a least-squares Monte-Carlo policy evaluation algorithm. Specifically, give a batch update equation that directly estimates θ, given multiple episodes of data.

[4 marks]

(f) Apply your least squares Monte-Carlo evaluation algorithm to estimate θ.

[4 marks]

(g) It would also be possible to solve this problem using least squares temporal-difference learning, LSTD(0). Would you expect your least squares Monte-Carlo estimate of θ to be better or worse than the LSTD estimate of θ? Explain your reasons. *Note:* do not compute the LSTD solution.

[4 marks]

[Total 20 marks]

6. A rat is in an experiment involving coloured light. Its state x is described by three continuous variables representing the colour of light: red (R), green (G) and blue (B), $x = \begin{bmatrix} R \\ G \\ B \end{bmatrix}$. At each time-step it can choose whether or not to press a button: $a = 1$ for pressing, and $a = 0$ for not pressing. The next colour of light depends on the current colour and whether or not the button is pressed. When the episode terminates, the rat receives some amount of food depending on the final colour of light. This problem is undiscounted ($\gamma = 1$).

The rat's policy is a logistic-linear distribution, $\pi_\theta(x,a) = \mathbb{P}[a|x] = \sigma(x^\top\theta)$ where σ is the logistic function $\sigma(x) = \frac{1}{1+e^{-x}}$, and θ is a vector of three parameters.

(a) The *score function* of a policy is defined by $\nabla_\theta \log \pi(x,a)$. What is the score function for the rat's policy? Hint: $\frac{d\sigma(y)}{dy} = \sigma(y)(1-\sigma(y))$. Also recall that $\frac{d \log x}{dx} = \frac{1}{x}$.

[4 marks]

(b) State the policy gradient theorem.

[2 marks]

(c) Use the policy gradient theorem and your score function to derive a REINFORCE algorithm for improving $\pi_\theta(x_t, a_t)$ by stochastic gradient ascent, after observing a return v_t from state x_t and action a_t. Your algorithm should be entirely policy-based (i.e. no value function).

[4 marks]

(d) At one particular time-step t, the rat is in a state $x_t = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$, with policy parameters $\theta = \begin{bmatrix} -1 \\ 0.5 \\ -0.5 \end{bmatrix}$. What is the probability of the rat pressing the button in this state, $\mathbb{P}[a_t = 1 \mid x_t]$?

[1 marks]

(e) In fact, at this time-step the rat presses the button, $a_t = 1$. At the end of the episode, the rat receives a reward of +2. Using a step-size of 0.1, apply your policy gradient update once for this time-step. What is the new parameter vector θ?

[3 marks]

A flea on the rat is observing the rat's behaviour. The flea estimates that pressing the button in state x_t has a value of +4 to the rat (and that not pressing the button has a value of -4).

(f) Now incorporate the flea's advice. Use the policy gradient theorem to derive an actor-critic algorithm for improving $\pi_\theta(x_t, a_t)$ by stochastic gradient ascent, when given an action-value $Q^\pi(s_t, a_t)$.

[3 marks]

(g) Assume again that the rat presses the button, $a_t = 1$. Using a step-size of 0.1, apply your new update once for this time-step. What is the new parameter vector θ?

[3 marks]

[Total 20 marks]

END OF PAPER