Remember to fill in the course feedback form
http://www.cs.hut.fi/Opinnot/Palaute/kurssipalaute-en.html

T-61.5040 Learning Models and Methods

Final exam 3.1.2007

You can use a non-programmable calculator and basic mathematics tables. Questions are only available in English, but you may answer in Finnish, Swedish, or English.

The exam has four problems: see also back of this paper!

1. (max 6p) Explain *briefly* the following concepts without unnecessary detail:
i) Posterior Distribution (2p)
ii) Laplace Approximation (2p)
iii) Jeffreys' Prior (2p)

2. (max 6p) You have observed (x, y) pairs $(0,0), (1,1), (2,2), (3,8), (4,4), (5,4)$. Your model for explaining y as a function of x is $y = x + c$. Assume that the likelihood $p(y|x, c)$ is Normal with mean $x + c$ and variance σ^2. The unknown quantities are c and the variance σ^2. Assume that $p(c)$ is constant and in part i) below, $p(\sigma)$ is constant. Find the *integer value* of c that maximizes the posterior when
i) The variance σ^2 is the same for all x; (2p)
ii) The variance σ^2 is drawn from $p(\sigma^2) \propto \sigma^{-7} \exp(-2\sigma^{-2})$ independently for all x_i. (4p)
Hint: Gamma-integral is $\Gamma(s) = \int_0^\infty z^{s-1} \exp(-z) dz$

3. (max 6p)

Consider a machine which can be diagnosed by taking independent measurements y_i. When the machine is working correctly, the measurements have a distribution $N(100, 0.2)$ (0.2 is the variance). When the machine has a fault, the measurement errors, i.e. $y_i - 100$, have a distribution $N(\mu, 0.4)$ where μ is unknown and has a prior $p(\mu) = N(\mu|0, 0.4)$. Assume that the machine is working correctly 95 percent of the time.
i) What is the predictive distribution for a new diagnostic measurement? (2p)
Hint: the iteration formulas are useful in this problem.

ii) Observe $y_1 = 102, y_2 = 99.9, y_3 = 101.5, y_4 = 100.1, y_5 = 98.7$. Compute the posterior for the variable L which is $L = 1$ if the machine is working correctly, and $L = 0$ if its not. (2p)
Hint: the density for a variable $z \sim N(0,1)$ *has values* $p(z = 2.2) \approx 0.035, p(z = 1.1) \approx 0.218$.

iii) Suppose that the only way to repair the machine is by stopping it first: this costs 100000 EUR regardless of the condition of the machine. If the machine is not working properly, it will cost 500000 EUR to keep operating it. Assume that operating a working machine is costless. Assuming you want to maximize the expected amount of money, should you stop the machine? (1p).

iv) If you can take five more observations, costing 150000 EUR in total, should you get the new observations? (1p)

4. (max 10p) The following statements are either true (T) or false (F). Write a table on your exam paper where the left column contains the statement numbers in ascending order and the right column your answers. Your answer to each statement must be an integer $q \in \{0, 1, 2, \ldots, 10\}$. The integer represents your belief as a probability (multiplied by ten) that the statement is *true*. You will receive $1 - 4(1 - 0.1 * q)^2$ points if the statement is true, and $1 - 4 * (0.1 * q)^2$ points if the statement is false. The scoring formulas are designed to give an incentive to actually answer your subjective probability, instead of e.g. guessing without knowing the answer. Note that the minimum number of points possible for one question is -3, and the average number of points obtained by guessing $q \in \{0, 10\}$ without information is -1. Answering $q = 5$ gives zero points. The maximum total score is the number of statements, and the minimum is zero (if you get a negative total score, it will be changed to zero).

1) If the prior is a Gamma-distribution $p(\theta) = G(\theta|a, b) \propto e^{-b\theta}\theta^{a-1}$ and the likelihood is a Poisson-distribution $p(y|\theta) = \frac{\theta^y e^{-\theta}}{y!}$, then the prior is not conjugate.
2) A model is selected optimally by computing the posterior distribution over the different models, and choosing the one with highest posterior probability.
3) Suppose an expert defines the model $p(D|\theta)$ for your problem and you believe that the model is correct. Then you use two different learning methods using the same data D. The learning method that results in less posterior uncertainty (e.g. smaller variance in $p(\theta|D)$ is always better.
4) In variational and free-form approximations the cost function to be minimized is the Kullback-Leibler divergence between the approximation $q(\theta)$ and the true posterior $p(\theta|y)$.
5) If a model contains nuisance parameters (not needed in the posterior), they are optimally handled by integrating the full posterior over all nuisance parameters.
6) If a hierarchical model contains observed data y and parameters a, b where b is a hyperparameter, then the posterior can be written as $p(a, b|y) \propto p(y|a)p(a|b)p(b)$.
7) Simulated samples generated by any Markov-Chain Monte Carlo method are independent.
8) Missing data is optimally handled by replacing missing values by estimated values.
9) **Points for this problem are doubled.** You participate in the tv-show "Who Wants to be a Millionaire", and you have just answered correctly to the 3000 EUR question. The rules of the game are such that you have the option to stop and receive the amount of money that your last correct answer is worth: in this case 3000 EUR. If you continue and give a wrong answer, you receive a smaller amount of money, in this case 2000 EUR. You are given four alternatives to each question, and exactly one of these is correct. The next question is worth 5000 EUR and the following question is worth 8000 EUR. Assume that you know nothing about the 5000 EUR question, i.e. you have to make a pure guess. Also assume that the 8000 EUR question is about Bayesian Inference, and therefore you decide that you can guess the correct answer with probability 0.75. Assume that you want to maximize the expected amount of money you receive. The statement of this problem is "Your optimal strategy is to stop the game and receive 3000 EUR".