Bayesian methods of Machine learning

Assignment 4

1. (Each bullet costs 1 point) There are three boxes, each contains black and white marbles. Their numbers are indicated in the table below. Five marbles are selected by selecting a box at random and then selecting a marble at random from that box with immediate replacement.

	box 1	box 2	box 3
black	2	3	1
white	2	1	3

- Assume the boxes selected are hidden. Formulate an HMM which models the above process for selecting marbles, i.e. determine $\lambda = (\pi, A, B)$, where π are initial hidden state probabilities, A is a matrix of state transition probabilities, B is a matrix of emission probabilities, depending on hidden states. Use 1 to represent black and 2 to represent white marbles.
- Compute the probability of observing the marble color sequence $O = (1, 1, 2, 2, 1)$ if the hidden sequence of boxes selected is $Q = (1, 1, 3, 3, 2)$.
- If the marble color sequence is $O = (1, 1, 2, 2, 1)$, determine the most likely sequence of boxes selected, Q^*. Explain how you obtained your answer. Hint: The fact that each of the hidden states are equally likely to have been selected at each step makes this an easy problem.
- Given the marble color sequence $O = (1, 1, 2, 2, 1)$, show that the answer to part (c), Q^*, is more likely than $Q = (1, 1, 3, 3, 2)$ from part (b) by computing a log-odds ratio. Hint: Compute $\frac{\log P(Q^*|O,\lambda)}{P(Q|O,\lambda)}$. You do not need to compute $P(O|\lambda)$.
- Compute $P(O|\lambda)$.

2. (4 points) A training sample $D = (X, \mathbf{y}) = \{\mathbf{x}_i, y_i = y(\mathbf{x}_i)\}_{i=1}^n$ of size n consists of values of a function $y(\mathbf{x}_i)$ at points $\mathbf{x}_i, i = \overline{1, n}$. We assume that the function $y(\mathbf{x})$ is a realization of a zero-mean Gaussian process with a covariance function $k_{\boldsymbol{\theta}}(\mathbf{x}, \tilde{\mathbf{x}})$, which depends on some parameters $\boldsymbol{\theta}$. We denote the sample with j-th element excluded as $D_{-j} = \{\mathbf{x}_i, y_i = y(\mathbf{x}_i)\}_{i=\overline{1,n}, i\neq j}$.

Let us denote by $\hat{y}_{-j} = \mathbb{E}(y(\mathbf{x}_j)|D_{-j}, \boldsymbol{\theta})$ the posterior mean of the Gaussian process at point $\mathbf{x}_j$ with a covariance function $k_{\boldsymbol{\theta}}(\mathbf{x}, \tilde{\mathbf{x}})$ given the sample D_{-j}. You are required to obtain an efficient approach for calculating leave-one-out cross validation squared error: $\sum_{j=1}^{n}(\hat{y}_{-j} - y_j)^2$, which does not require re-calculating value of $\hat{y}_{-j}$ for each $j = 1, \ldots, n$.

3. (Each bullet costs 2 points) We assume that the function $y(\mathbf{x})$ is a realization of a zero-mean Gaussian process with a sufficiently smooth covariance function $\operatorname{cov}(y(\mathbf{x}), y(\tilde{\mathbf{x}})) = k_{\boldsymbol{\theta}}(\mathbf{x}, \tilde{\mathbf{x}})$, which depends on some parameters $\boldsymbol{\theta}$. Suppose, that we have two samples of observations: $D = (X, \mathbf{y}) = \{\mathbf{x}_i, y_i = y(\mathbf{x}_i)\}_{i=1}^{n}$ and $D_1 = (X', \mathbf{y}_1) = \left\{\mathbf{x}'_i, y_{i1} = \frac{\partial y(\mathbf{x})}{\partial x_1}\big|_{\mathbf{x}=\mathbf{x}'_i}\right\}_{i=1}^{m}$.

- Calculate a covariance function $\operatorname{cov}\left(y(\mathbf{x}), \frac{\partial y(\mathbf{x})}{\partial x_1}\big|_{\mathbf{x}=\tilde{\mathbf{x}}}\right)$
- Find the mean and the variance for the conditional distribution of $y(\mathbf{x})|D, D_1$.
- Prove, that the conditional distribution of $y(\mathbf{x})|D, D_1$ is Gaussian.