ANSWERS FOR THIRD TEST

1/20	
2/10	
3/15	
4/19	
5/36	
T/100	

STA 414/2104 — Third Test — 2007-04-13

No books, notes, or calculators are allowed. Time is 50 minutes.

Recall that Gaussian process regression for a response t given inputs x is based on a model in which the response for a case is modeled as $t = y(x) + \epsilon$, with $\epsilon \sim N(0, \beta^{-1})$ and $y(x)$ having a Gaussian process prior distribution with mean zero and $\text{Cov}(y(x), y(x')) = k(x, x')$, for some function $k(\cdot, \cdot)$. In this framework, the predictive distribution for the response t^* in a test case with inputs x^* has mean and variance given by

$$E[t^* \,|\, x^*, \text{training data}] \;=\; k^T C^{-1} t$$
$$\text{Var}[t^* \,|\, x^*, \text{training data}] \;=\; c \,-\, k^T C^{-1} k$$

where t is the vector of observed responses in training cases, C is the matrix of covariances for the responses in training cases, k is the vector of covariances of the response in the test case with the responses in training cases, and c is the prior variance of the response in the test case.

For Questions 1 to 3 below, x is one-dimensional.

Question 1: [20 marks] Suppose we have just one training case, with $x_1 = 3$ and $t_1 = 4$. Suppose also that $\beta = 1$ and that $k(x, x') = 2^{-|x-x'|}$. Find the mean and variance of the predictive distribution for the response in a test case for which the value of the input is 5.

With just one training case, C is a 1 by 1 matrix, and can be treated as a scalar. Similarly for k, a vector of length 1. C is the sum of the prior variance for the training case, which is $k(3, 3) = 1$, and the noise variance, which is $\beta^{-1} = 1$. k is equal to the covariance of the training case and test case, which is $k(3, 5) = 1/4$. The prior variance of the test case is $c = k(5, 5) + \beta^{-1} = 2$. The predictive mean and variance are therefore

$$E[t^* \,|\, x^*, \textit{training data}] \;=\; (1/4)(1/2)4 \;=\; 1/2$$
$$Var[t^* \,|\, x^*, \textit{training data}] \;=\; 2 \,-\, (1/4)(1/2)(1/4) \;=\; 63/32$$

Question 2: [10 marks] Repeat the calculations Question 2, but using $k(x, x') = 2^{+|x-x'|}$. What can you conclude from the result of this calculation?

With this definition of $k(x, x')$:

$$E[t^* \,|\, x^*, \textit{training data}] \;=\; 4(1/2)4 \;=\; 8$$
$$Var[t^* \,|\, x^*, \textit{training data}] \;=\; 2 \,-\, 4(1/2)4 \;=\; -6$$

Since negative variances are impossible, we can conclude that $k(x, x') = 2^{+|x-x'|}$ is not a valid covariance function.

Question 3: [15 marks] Suppose we have two training cases, with $x_1 = 3.2$, $t_1 = -2$, $x_2 = 4.4$, and $t_2 = 3$. Suppose also that $\beta = 1$ and that we use the following covariance function:

$$k(x, x') = \begin{cases} 1 - |x - x'| & \text{if } |x - x'| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the mean and variance of the predictive distribution for the response in a test case for which the value of the input is 4.

Since $|x_1 - x_2| = 1.2 > 1$, the matrix C will be diagonal, so matrix inversion is easy. We get

$$E[t^* \,|\, x^*, \textit{training data}] = [\,0.2 \;\; 0.6\,] \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}^{-1} \begin{bmatrix} -2 \\ 3 \end{bmatrix} = 0.7$$

$$Var[t^* \,|\, x^*, \textit{training data}] = 2 - [\,0.2 \;\; 0.6\,] \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 0.2 \\ 0.6 \end{bmatrix} = 1.8$$

Question 4: [19 marks] Consider a Bayesian linear regression model with one input, in which the regression function has the form

$$y(x) = w_0 + w_1 I(x > 0) + w_2 I(x \leq 0)$$

where $I(\cdot)$ is the indicator function, equal to 1 if the condition inside is true, and 0 if not. Suppose that in our prior distribution, w_0, w_1, and w_2 are independent, with $w_0 \sim N(0, 2^2)$, $w_1 \sim N(0, 3^2)$, and $w_2 \sim N(0, 4^2)$. Find the covariance function, $k(x, x') = \text{Cov}(y(x), y(x'))$ for the equivalent Gaussian process model.

Since the means are zero, the covariance is just the expectation of the product. Also, note that any term like $E[w_1 w_2 F]$ will be zero, since w_1 and w_2 are independent and have means of zero. We can therefore find the covariance function as follows:

$$\begin{aligned}
k(x, x') &= E[y(x)y(x')] \\
&= E[\,(w_0 + w_1 I(x > 0) + w_2 I(x \leq 0))\,(w_0 + w_1 I(x' > 0) + w_2 I(x' \leq 0))\,] \\
&= E[\,w_0^2 + w_0 w_1 I(x' > 0) + w_0 w_2 I(x' \leq 0) \\
&\qquad + w_1 w_0 I(x > 0) + w_1^2 I(x > 0) I(x' > 0) + w_1 w_2 I(x > 0) I(x' \leq 0) \\
&\qquad + w_2 w_0 I(x \leq 0) + w_1 w_2 I(x \leq 0) I(x' > 0) + w_2^2 I(x \leq 0) I(x' \leq 0)\,] \\
&= E[w_0^2] + E[w_1^2 I(x > 0) I(x' > 0)] + E[w_2^2 I(x \leq 0) I(x' \leq 0)] \\
&= 4 + 9 I(x > 0) I(x' > 0) + 16 I(x \leq 0) I(x' \leq 0) \\
&= \begin{cases} 20 & \text{if } x \leq 0 \text{ and } x' \leq 0 \\ 13 & \text{if } x > 0 \text{ and } x' > 0 \\ 4 & \text{otherwise} \end{cases}
\end{aligned}$$

Question 5: [36 marks] Consider the following four covariance functions for a Gaussian process with a two-dimensional input vector, $x = (x_1, x_2)$:

(a) $k(x, x') \;=\; 1 \;+\; 5^2 \exp(-(x_1 - x_1')^2/5^2) \;+\; 0.2^2 \exp(-(x_2 - x_2')^2/0.25^2)$

(b) $k(x, x') \;=\; 1 \;+\; x_1 x_1' \;+\; 0.2^2 \exp(-(x_2 - x_2')^2)$

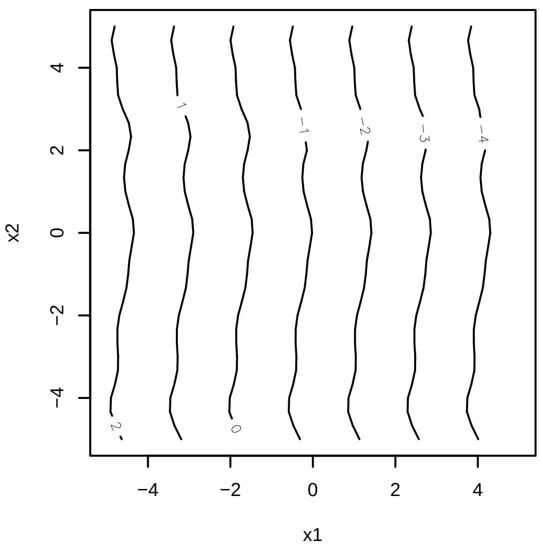
(c) $k(x, x') \;=\; 1 \;+\; 5^2 \exp(-(x_1 - x_1')^2 - (x_2 - x_2')^2/5^2)$

(d) $k(x, x') \;=\; 1 \;+\; 5^2 \exp(-(x_1 - x_1')^2/2^2 - (x_2 - x_2')^2/2^2)$

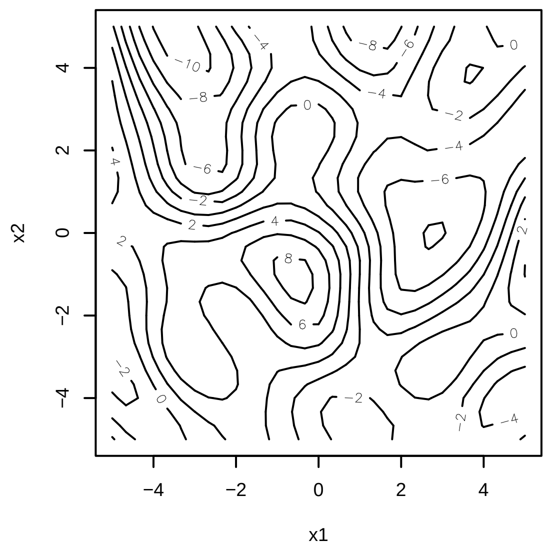
(e) $k(x, x') \;=\; 1 \;+\; 5^2 \exp(-(x_1 - x_1')^2/10^2) \;+\; 5^2 \exp(-(x_2 - x_2')^2)$

Each of the four plots below is of a random sample from a Gaussian process with one of these covariance functions. For each plot, identify which covariance function was used to generate it. A covariance function might (or might not) be used more than once.

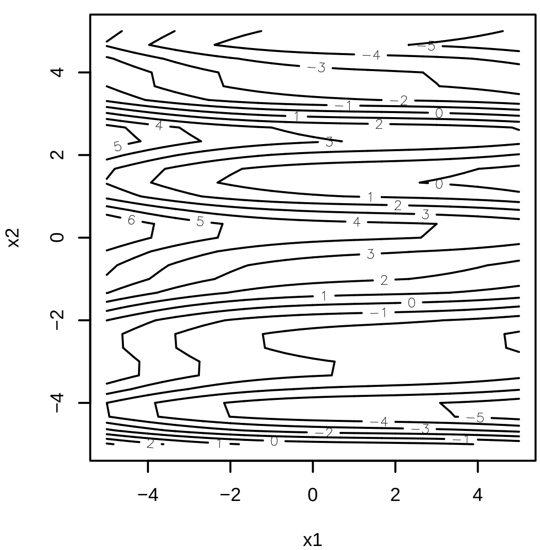
Covariance function: b



Covariance function: d



Covariance function: e



Covariance function: a