STA 414/2104, Spring 2013, Practice Problem Set #2, Answers

Question 1: Suppose we model the relationship of a real-valued response variable, y, to a single real input, x, using a Gaussian process model in which the mean is zero and the covariances of the observed responses are given by

$$\text{Cov}(y_i, y_{i'}) \;=\; 0.5^2\delta_{i,i'} \;+\; K(x_i, x_{i'})$$

with the noise-free covariance function, K, defined by

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

Suppose we have four training cases, as follows:

x	y
0.5	2.0
2.8	3.3
1.6	3.0
3.9	2.7

Recall that the conditional mean of the response in a test case with input x_*, given the responses in the training cases, is $k^TC^{-1}y$, where y is the vector of training responses, C is the covariance matrix of training responses, and k is the vector of convariances of training responses with the response in the test case.

Find the predictive mean for the response in a test case in which the input is $x_* = 1.2$.

The covariance matrix of the training responses is

$$C \;=\; \begin{bmatrix} 1+0.5^2 & 0 & 0 & 0 \\ 0 & 1+0.5^2 & 0 & 0 \\ 0 & 0 & 1+0.5^2 & 0 \\ 0 & 0 & 0 & 1+0.5^2 \end{bmatrix}$$

The inverse of this is

$$C^{-1} \;=\; \begin{bmatrix} 0.8 & 0 & 0 & 0 \\ 0 & 0.8 & 0 & 0 \\ 0 & 0 & 0.8 & 0 \\ 0 & 0 & 0 & 0.8 \end{bmatrix}$$

The vector of covariances of the test response with the training responses is

$$k \;=\; \begin{bmatrix} 1-0.7 \\ 0 \\ 1-0.4 \\ 0 \end{bmatrix} \;=\; \begin{bmatrix} 0.3 \\ 0 \\ 0.6 \\ 0 \end{bmatrix}$$

So $k^TC^{-1} \;=\; [0.24\ 0\ 0.48\ 0]$, and the predictive mean for the test resp[onse is

$$k^TC^{-1}y \;=\; 0.24 \times 2.0 \;+\; 0.48 \times 3.0 \;=\; 1.92$$

Question 2: Recall that for a Gaussian process model the predictive distribution for the response y^* in a test case with inputs x^* has mean and variance given by

$$\begin{aligned} E[y^* \,|\, x^*, \text{training data}] &= k^T C^{-1} y \\ \text{Var}[y^* \,|\, x^*, \text{training data}] &= v - k^T C^{-1} k \end{aligned}$$

where y 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 v is the prior variance of the response in the test case.

a) Suppose we have just one training case, with $x_1 = 3$ and $y_1 = 4$. Suppose also that the noise-free covariance function is $K(x, x') = 2^{-|x-x'|}$, and the variance of the noise is $1/2$. 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.

The mean of the preditive distribution is

$$K(3,5)[K(3,3) + 1/2]^{-1}(4) = (1/4)[1 + 1/2]^{-1}(4) = 4/6$$

The variance of the predictive distribution is

$$[K(5,5)+1/2]-K(3,5)[K(3,3)+1/2]^{-1}K(3,5) = [1+1/2]-(1/4)[1+1/2]^{-1}(1/4) = 35/24$$

b) Repeat the calculations for (a), but using $K(x, x') = 2^{+|x-x'|}$. What can you conclude from the result of this calculation?

The mean of the preditive distribution is

$$K(3,5)[K(3,3) + 1/2]^{-1}(4) = (4)[1 + 1/2]^{-1}(4) = 32/3$$

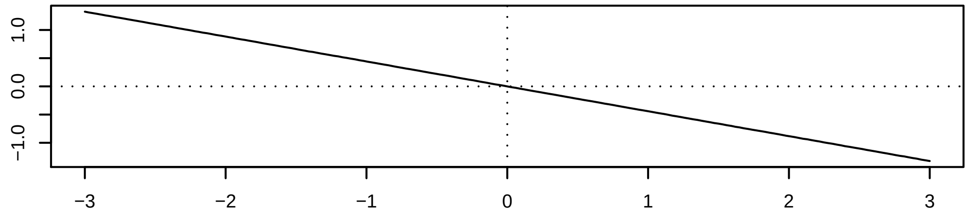
The variance of the predictive distribution is

$$[K(5,5)+1/2]-K(3,5)[K(3,3)+1/2]^{-1}K(3,5) = [1+1/2]-(4)[1+1/2]^{-1}(4) = -55/6$$

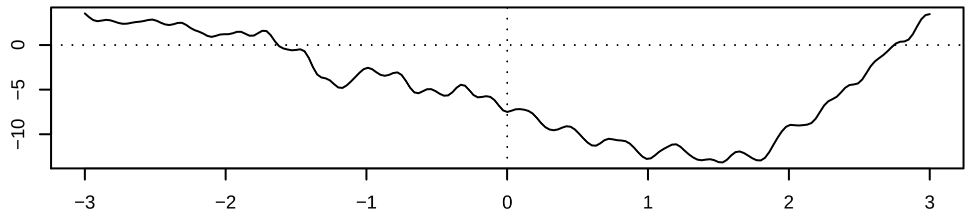
But variances cannot be negative! We can conclude that $K(x, x') = 2^{+|x-x'|}$ *is not a valid covariance function — it is not positive semi-definite.*

Question 3: Below are five functions randomly drawn from five different Gaussian processes. For all five Gaussian processes, the mean function is zero. The covariance functions are one of those listed below.

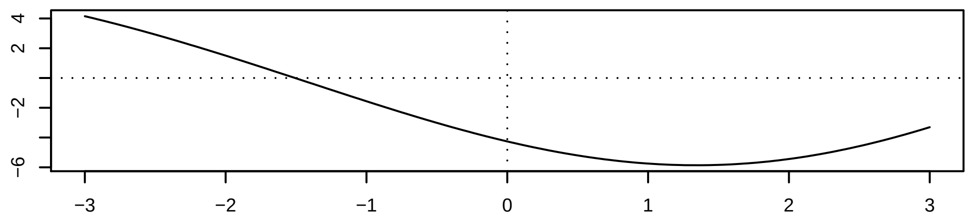
a)



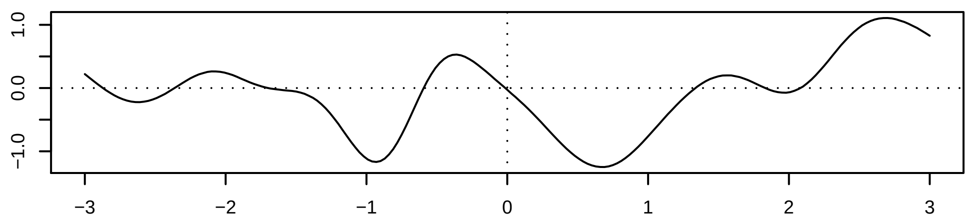
b)



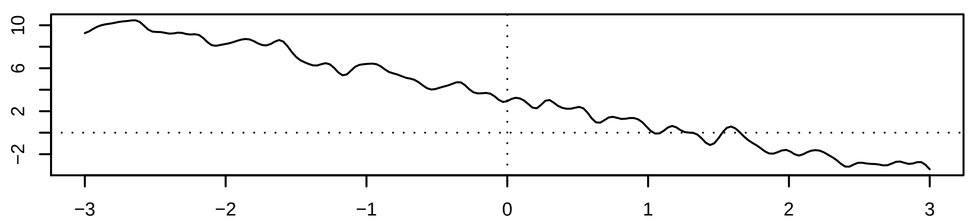
c)



d)



e)



For each of the five covariance functions below, indicate which of the five functions above is most likely to have been drawn from the Gaussian process with that covariance function.

1) $\text{Cov}(y_{i_1}, y_{i_2}) \;=\; 0.5^2 \exp(-((x_{i_1} - x_{i_2})\,/\,0.5)^2)$ *Answer: (d)*

2) $\text{Cov}(y_{i_1}, y_{i_2}) \;=\; x_{i_1}\, x_{i_2}$ *Answer: (a)*

3) $\text{Cov}(y_{i_1}, y_{i_2}) \;=\; 5^2 \;+\; 5^2\, x_{i_1}\, x_{i_2} \;+\; 0.5^2 \exp(-((x_{i_1} - x_{i_2})\,/\,0.1)^2)$ *Answer: (e)*

4) $\text{Cov}(y_{i_1}, y_{i_2}) \;=\; 0.7^2 \exp(-((x_{i_1} - x_{i_2})\,/\,0.1)^2) \;+\; 8^2 \exp(-((x_{i_1} - x_{i_2})\,/\,2)^2)$ *Answer: (b)*

5) $\text{Cov}(y_{i_1}, y_{i_2}) \;=\; 8^2 \exp(-((x_{i_1} - x_{i_2})\,/\,5)^2)$ *Answer: (c)*

Question 4: Suppose that we are fitting a Gaussian mixture model for data items consisting of a single real value, x, using $K = 2$ components. We have $N = 5$ training cases, in which the values of x are as follows:

$$5, \quad 15, \quad 25, \quad 30, \quad 40$$

We use the EM algorithm to find the maximum likeihood estimates for the model parameters, which are the mixing proportions for the two components, π_1 and π_2, and the means for the two components, μ_1 and μ_2. The standard deviations for the two components are fixed at 10.

Suppose that at some point in the EM algorithm, the E step found that the responsibilities of the two components for the five data items were as follows:

r_{i1}	r_{i2}
0.2	0.8
0.2	0.8
0.8	0.2
0.9	0.1
0.9	0.1

What values for the parameters π_1, π_2, μ_1, and μ_2 will be found in the next M step of the algorithm?

The new estimates will be

$$\pi_1 = (0.2 + 0.2 + 0.8 + 0.9 + 0.9)/5 = 0.6$$

$$\pi_2 = (0.8 + 0.8 + 0.2 + 0.1 + 0.1)/5 = 0.4$$

$$\mu_1 = (0.2 \times 5 + 0.2 \times 15 + 0.8 \times 25 + 0.9 \times 30 + 0.9 \times 40) \,/\, (0.2 + 0.2 + 0.8 + 0.9 + 0.9) = 29$$

$$\mu_2 = (0.8 \times 5 + 0.8 \times 15 + 0.2 \times 25 + 0.1 \times 30 + 0.1 \times 40) \,/\, (0.8 + 0.8 + 0.2 + 0.1 + 0.1) = 14$$

Question 5: Consider a two-component Gaussian mixture model for univariate data, in which the probability density for an observation, x, is

$$(1/2)N(x|\mu, 1) \;+\; (1/2)N(x|\mu, 2^2)$$

Here, $N(x|\mu, \sigma^2)$ denotes the density for x under a univariate normal distribution with mean μ and variance σ^2. Notice that mixing proportions are equal for this mixture model, that the two components have the same mean, and that the standard deviations of the two components are fixed at 1 and 2. There is only one model parameter, μ.

Suppose we wish to estimate the μ parameter by maximum likelihood using the EM algorithm. Answer the following questions regarding how the E step and M step of this algorithm operate, if we have the three data points below:

$$4.0, \;\; 4.6, \;\; 2.0$$

Here is a table of standard normal probability densities that you may find useful:

x	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
$N(x\|0,1)$	.40	.40	.39	.38	.37	.35	.33	.31	.29	.27	.24	.22	.19	.17	.15	.13	.11	.09	.08	.07	.05

a) Find the responsibilities that will be computed in the E step if the model parameter estimates from the previous M step are $\mu = 4$, $\sigma_1 = 1$, and $\sigma_2 = 2$. Since the responsibilities for the two components must add to one, it is enough to give $r_{i1} = P(\text{component } 1 \,|\, x_i)$ for $i = 1, 2, 3$.

First, note that the normal density function with mean μ *and variance* σ^2 *is* $N(x|\mu,\sigma^2) = (1/\sigma)N((x-\mu)/\sigma|0,1)$. *Also* $N(-x|0,1) = N(x|0,1)$.

Using Bayes' Rule, we get that

$$P(\text{component } 1|x) \quad = \quad \frac{(1/2)N(x|\mu,1)}{(1/2)N(x|\mu,1) + (1/2)N(x|\mu,2^2)}$$

Applying this the three observations, we get

$$\begin{aligned} r_1 1 &= \frac{(1/2)0.40}{(1/2)0.40 + (1/2)(1/2)0.40} &= 2/3 \\ r_2 1 &= \frac{(1/2)0.33}{(1/2)0.33 + (1/2)(1/2)0.38} &= 33/52 \\ r_3 1 &= \frac{(1/2)0.05}{(1/2)0.05 + (1/2)(1/2)0.24} &= 5/17 \end{aligned}$$

b) Using the responsibilities that you computed in part (a), find the estimate for μ that will be found in the next M step. Recall that the M step maximizes the expected value of the log of the probability density for x_1, x_2, x_3 and the unknown component indicators, with the expectation taken with respect to the distribution for the component indicators found in the previous E step.

The expected log likelihood is

$$\sum_{i=1}^{3} \left[r_{i1}(-(1/2)(x_i-\mu)^2) \ + \ (1-r_{i1})(-(1/2)(x_i-\mu)/2^2) \right]$$

To find the maximum of this with respect to μ, *we take the derivative with respect to* μ, *which is*

$$\sum_{i=1}^{3} [r_{i1}(x_i-\mu) \ + \ (1-r_{i1})(x_i-\mu)/4]$$

Setting this to zero and solving for μ *gives*

$$\hat{\mu} \quad = \quad \frac{\sum_{i=1}^{3}(r_{i1} + (1-r_{i1})/4)\, x_i}{\sum_{i=1}^{3}(r_{i1} + (1-r_{i1})/4)} \quad = \quad \frac{(3/4)4.0 + (151/208)4.6 + (25/68)2.0}{(3/4) + (151/208) + (25/68)}$$