

Lecture 5: Gaussian Process Covariance Functions

Machine Learning 4F13, Michaelmas 2015

Zoubin Ghahramani

Department of Engineering
University of Cambridge

<http://mlg.eng.cam.ac.uk/teaching/4f13/>

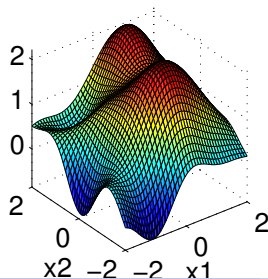
Model Selection, Hyperparameters, and ARD

We need to determine both the *form* and *parameters* of the covariance function. We typically use a **hierarchical model**, where the parameters of the covariance are called **hyperparameters**.

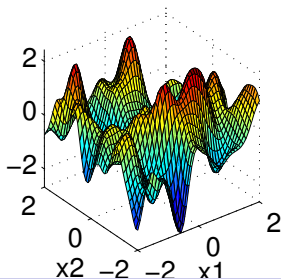
A very useful idea is to use **automatic relevance determination (ARD)** covariance functions for feature/variable selection, e.g.:

$$k(\mathbf{x}, \mathbf{x}') = v_0^2 \exp\left(-\sum_{d=1}^D \frac{(x_d - x'_d)^2}{2v_d^2}\right), \quad \text{hyperparameters } \theta = (v_0, v_1, \dots, v_d, \sigma_n^2).$$

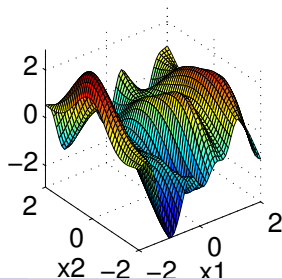
$v_1=v_2=1$



$v_1=v_2=0.32$



$v_1=0.32$ and $v_2=1$



Rational quadratic covariance function

The *rational quadratic* (RQ) covariance function, where $r = x - x'$:

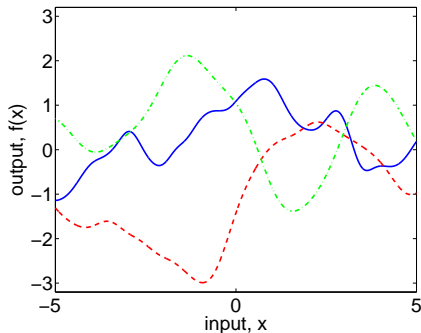
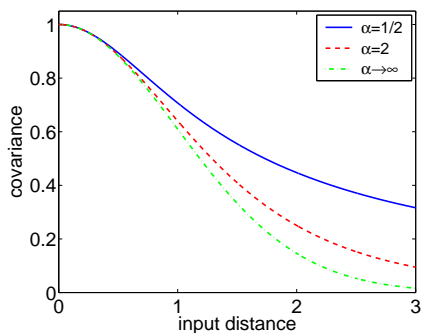
$$k_{\text{RQ}}(r) = \left(1 + \frac{r^2}{2\alpha\ell^2}\right)^{-\alpha}$$

with $\alpha, \ell > 0$ can be seen as a *scale mixture* (an infinite sum) of squared exponential (SE) covariance functions with different characteristic length-scales.

Using $\tau = \ell^{-2}$ and $p(\tau|\alpha, \beta) \propto \tau^{\alpha-1} \exp(-\alpha\tau/\beta)$:

$$\begin{aligned}k_{\text{RQ}}(r) &= \int p(\tau|\alpha, \beta) k_{\text{SE}}(r|\tau) d\tau \\ &\propto \int \tau^{\alpha-1} \exp\left(-\frac{\alpha\tau}{\beta}\right) \exp\left(-\frac{\tau r^2}{2}\right) d\tau \propto \left(1 + \frac{r^2}{2\alpha\ell^2}\right)^{-\alpha},\end{aligned}$$

Rational quadratic covariance function II



The limit $\alpha \rightarrow \infty$ of the RQ covariance function is the SE.

Matérn covariance functions

Stationary covariance functions can be based on the Matérn form:

$$k(\mathbf{x}, \mathbf{x}') = \frac{1}{\Gamma(\nu)2^{\nu-1}} \left[\frac{\sqrt{2\nu}}{\ell} |\mathbf{x} - \mathbf{x}'| \right]^\nu K_\nu \left(\frac{\sqrt{2\nu}}{\ell} |\mathbf{x} - \mathbf{x}'| \right),$$

where K_ν is the modified Bessel function of second kind of order ν , and ℓ is the characteristic length scale.

Sample functions from Matérn forms are $\lfloor \nu - 1 \rfloor$ times differentiable. Thus, the hyperparameter ν can control the degree of smoothness

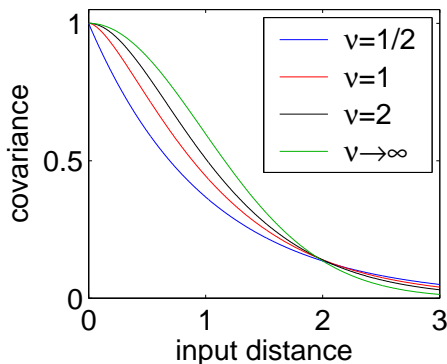
Special cases:

- $k_{\nu=1/2}(r) = \exp(-\frac{r}{\ell})$: Laplacian covariance function, Brownian motion (Ornstein-Uhlenbeck)
- $k_{\nu=3/2}(r) = (1 + \frac{\sqrt{3}r}{\ell}) \exp(-\frac{\sqrt{3}r}{\ell})$ (once differentiable)
- $k_{\nu=5/2}(r) = (1 + \frac{\sqrt{5}r}{\ell} + \frac{5r^2}{3\ell^2}) \exp(-\frac{\sqrt{5}r}{\ell})$ (twice differentiable)
- $k_{\nu \rightarrow \infty} = \exp(-\frac{r^2}{2\ell^2})$: smooth (infinitely differentiable)

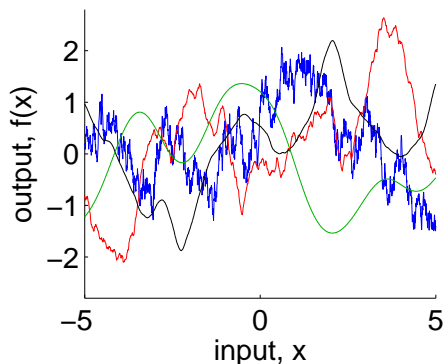
Matérn covariance functions II

Univariate Matérn covariance function with unit characteristic length scale and unit variance:

covariance function



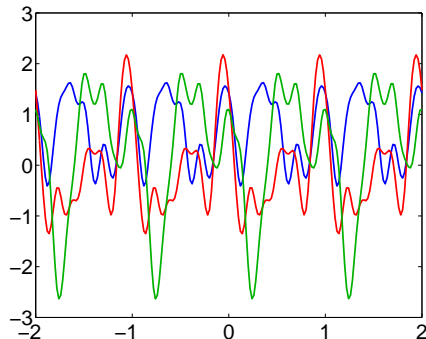
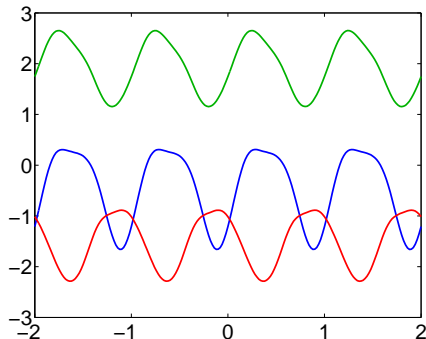
sample functions



Periodic, smooth functions

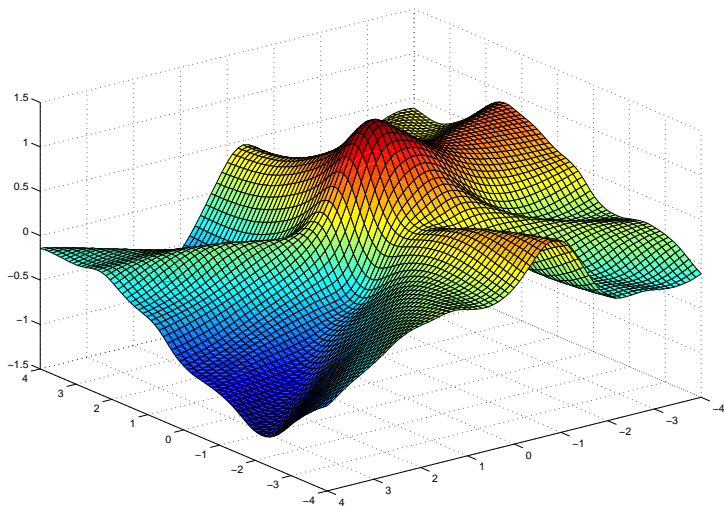
To create a distribution over periodic functions of x , we can first map the inputs to $u = (\sin(x), \cos(x))^T$, and then measure distances in the u space. Combined with the SE covariance function, which characteristic length scale ℓ , we get:

$$k_{\text{periodic}}(x, x') = \exp(-2 \sin^2(\pi(x - x'))/\ell^2)$$



Three functions drawn at random; left $\ell > 1$, and right $\ell < 1$.

Function drawn at random from a Neural Network covariance function



$$k(x, x') = \frac{2}{\pi} \arcsin \left(\frac{2x^\top \Sigma x'}{\sqrt{(1 + x^\top \Sigma x)(1 + 2x'^\top \Sigma x')}} \right).$$