Lecture 5: Gaussian Process Covariance Functions

Machine Learning 4F13, Spring 2014

Carl Edward Rasmussen and Zoubin Ghahramani

CUED

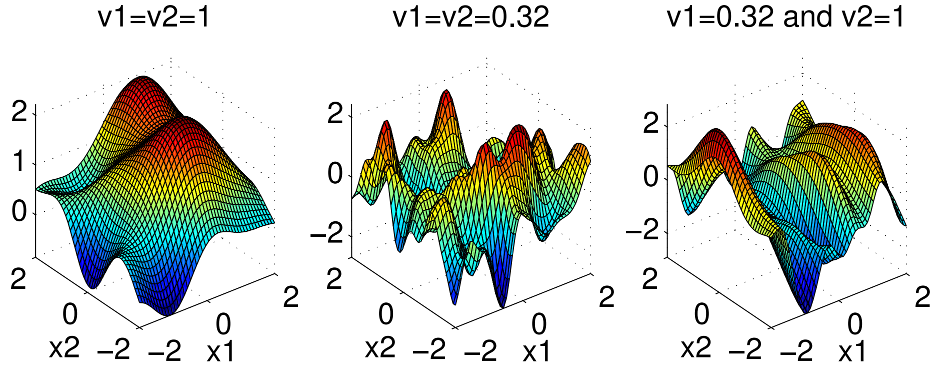
`http://mlg.eng.cam.ac.uk/teaching/4f13/`

Model Selection in Practice; Hyperparameters

There are two types of task: *form* and *parameters* of the covariance function.

Typically, our prior is too weak to quantify aspects of the covariance function. We use a hierarchical model using hyperparameters. Eg, in ARD:

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



Rational quadratic covariance function

The *rational quadratic* (RQ) covariance function:

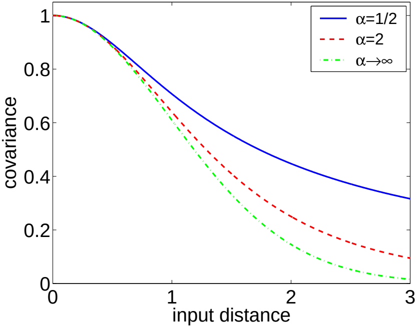
$$k_{RQ}(r) = \left(1 + \frac{r^2}{2\alpha\ell^2}\right)^{-\alpha}$$

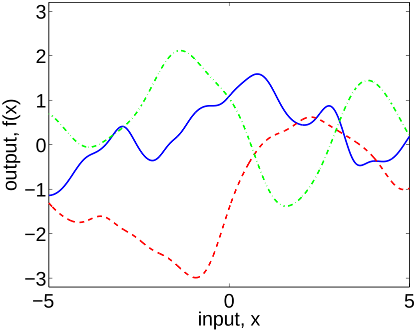
with $\alpha,\ \ell > 0$ can be seen as a *scale mixture* (an infinite sum) of squared exponential (SE) covariance functions with different characteristic length-scales.

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

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

Rational quadratic covariance function II





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

Matérn covariance functions

Stationary covariance functions can be based on the Matérn form:

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

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

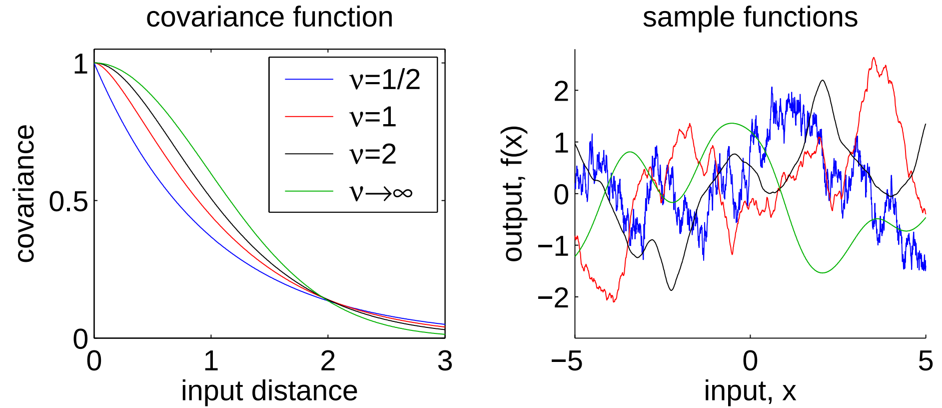
Sample functions from Matérn forms are $\lfloor \nu - 1 \rfloor$ times differentiable. Thus, the hyperparameter ν can control the degree of smoothness

Special cases:

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

Matérn covariance functions II

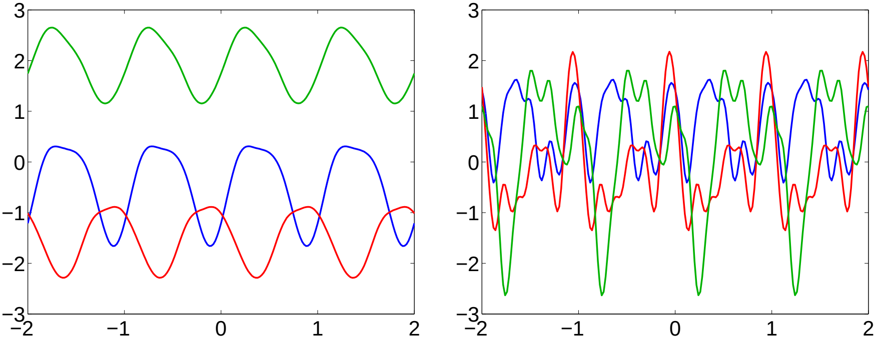
Univariate Matérn covariance function with unit characteristic length scale and unit variance:



Periodic, smooth functions

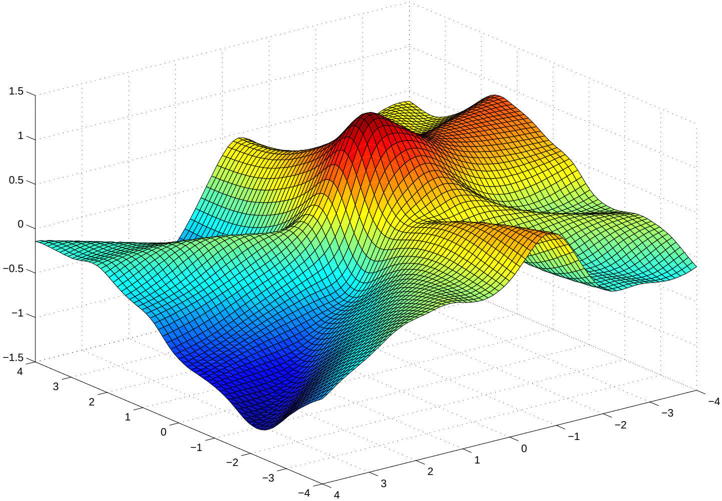
To create a distribution over periodic functions of x, we can first map the inputs to $u = (\sin(x), \cos(x))^\top$, and then measure distances in the u space. Combined with the SE covariance function, which characteristic length scale ℓ, we get:

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



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

Function drawn at random from a Neural Network covariance function



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