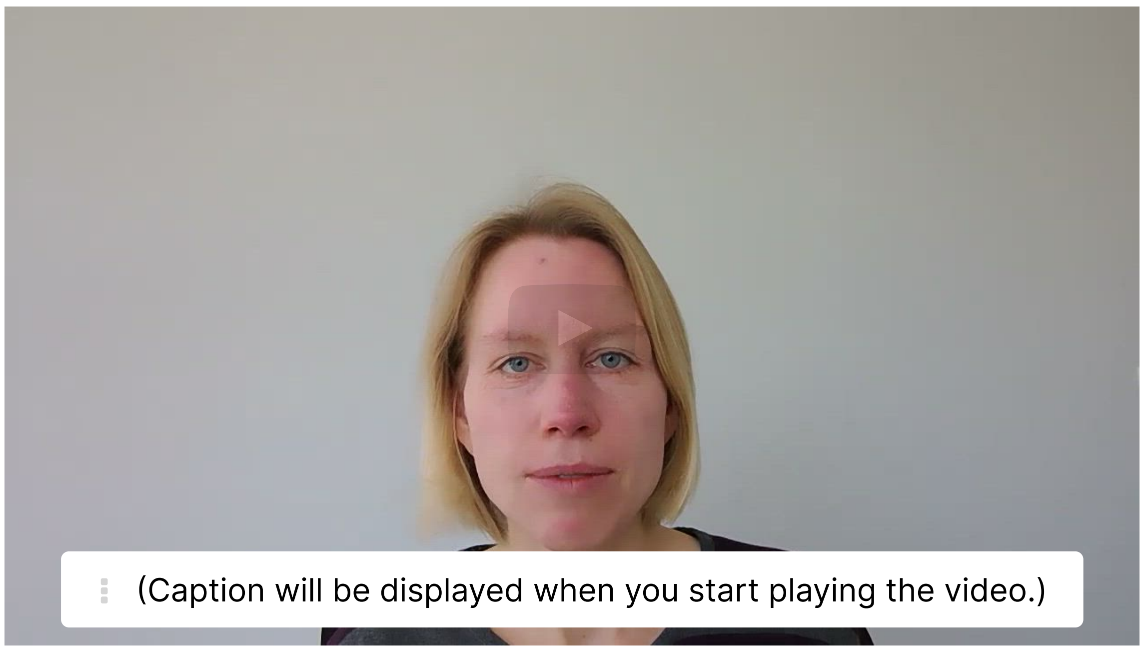
5. Effects of Parameters on Kernel Functions

Bookmark this page

Exercises due Dec 1, 2021 17:29 IST

Effects of Parameters on Kernel Functions



0:00 / 0:00 2.0x

Start of transcript. Skip to the end.

Prof Jegelka: So for the rest of the lecture,

I would just like to take you through a journey

through many different of these covariant functions

and show you what the effects are.

And you can actually see it's very interesting and rich,

the kinds of pattern you can see.

Video

Download video file

Transcripts

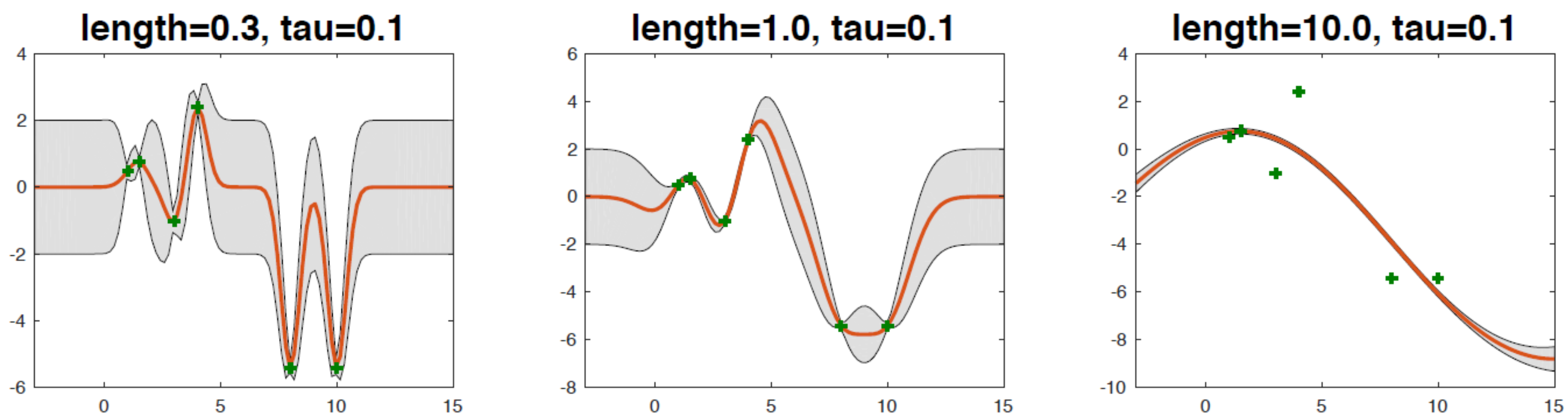
Download SubRip (.srt) file

Download Text (.txt) file

The below figure shows the effects of selecting different values of ℓ in the selected kernel function. In the figure, we observe the curve estimation, as shown in the previous example, with six observations. We see the effects of different values of the selected parameter ℓ: as one selects a larger ℓ the interpolations become smoother.

Note that higher values do not necessarily mean better estimates. This is a parameter that needs to be selected carefully. For example, in the left-most image, a value that is too small introduces artifacts in the estimates that might not exist in reality, see, for instance, the region between the right-most two points, where the predicted mean departs sharply back to zero. A value that is too large is shown in the right-most image; this ignores, or deletes, some of the changes observed in the data leading to a prediction that is too smooth.

$$k(x_i, x_j) = \exp\left(-\frac{||x_i - x_j||^2}{2\ell^2}\right)$$



kernel function determines shape of interpolation

33: Effects of the parametrization of a kernel function

Note also that the same ℓ parameter does not need to be used for each coordinate of the observations, one can have kernel functions with different parameters for each coordinate.

The below table shows other examples of possible forms for kernel functions. Assume here that $\boldsymbol{x}$ is a difference between points, eg: $\boldsymbol{x} = Z_1 - Z_2$, and $\boldsymbol{r}$ is a distance, eg: $\boldsymbol{r} = ||Z_1 - Z_2||$

between points, eg. $\boldsymbol{x} = Z_1 - Z_2$, and r is a distance, eg. $r = \|Z_1 - Z_2\|$.

constant	σ_0^2
linear	$\sum_{d=1}^{D} \sigma_d^2 x_d x_d'$
polynomial	$(\boldsymbol{x} \cdot \boldsymbol{x}' + \sigma_0^2)^p$
squared exponential	$\exp\left(-\frac{r^2}{2\ell^2}\right)$
Matérn	$\frac{1}{2^{\nu-1}\Gamma(\nu)}\left(\frac{\sqrt{2\nu}}{\ell}r\right)^{\nu} K_{\nu}\left(\frac{\sqrt{2\nu}}{\ell}r\right)$
exponential	$\exp\left(-\frac{r}{\ell}\right)$
γ-exponential	$\exp\left(-\left[\frac{r}{\ell}\right]^{\gamma}\right)$
rational quadratic	$\left(1+\frac{r^2}{2\alpha\ell^2}\right)^{-\alpha}$
neural network	$\sin^{-1}\left(\frac{2\boldsymbol{x}^{\top}\boldsymbol{\Sigma}\boldsymbol{x}}{\sqrt{(1+2\boldsymbol{x}^{\top}\boldsymbol{\Sigma}\boldsymbol{x})(1+2\boldsymbol{x}'^{\top}\boldsymbol{\Sigma}\boldsymbol{x}')}}\right)$

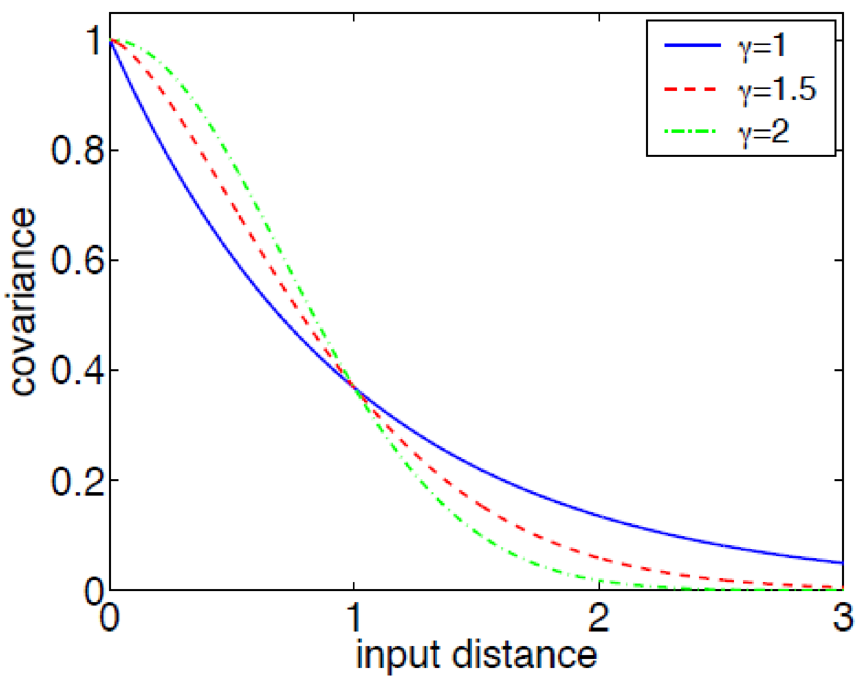
Note, for example, that the last kernel in this table is stationary, but it is not isotropic as the matrix $\boldsymbol{\Sigma}$ means that the kernel is no longer a function of purely distance. Rather, the distance between two points is scaled depending on the direction of the vector between them; hence, the kernel is not isotropic.

For example, consider the following covariance function:

$$k(Z_1, Z_2) = \exp\left(-\left[\frac{\|Z_1 - Z_2\|}{\ell}\right]^{\gamma}\right).$$

This is the γ-exponential shown in table above.

Examples of this function are shown below for various choices of γ.



34: The effects of the parameter γ on the covariance function.
When $\gamma = 1$, the peak of the function is sharp, and the function is heavy tailed. For $\gamma = 2$ the function is smooth around the peak, but falls off sharply at large distance. Although $\gamma = 2$ is a common choice, we can make other choices if we have reason to suspect that the observations are heavily correlated at large distance.

We can also combine kernel functions to form new kernel functions.

If k_1 is a kernel function, and k_2 is also a kernel function, then

$$k\left(Z_1, Z_2\right)=k_1\left(Z_1, Z_2\right)+k_2\left(Z_1, Z_2\right)$$

is also a kernel function.

Similarly,

$$k\left(Z_1, Z_2\right)=k_1\left(Z_1, Z_2\right) \times k_2\left(Z_1, Z_2\right)$$

is also a kernel function.

Combinations of kernel functions 1

1/1 point (graded)
Suppose that k_1 and k_2 are both isotropic kernel functions.

Is the linear combination

$$k\left(Z_1, Z_2\right)=2k_1\left(Z_1, Z_2\right)+k_2\left(Z_1, Z_2\right)$$

also an isotropic kernel function?

- (selected) Yes
- No

✔

Solution:

If k_1 and k_2 are isotropic, then they only depend on the distance between the supplied points. Therefore k also only depends on this distance, and nothing else. Thus k is also isotropic.

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Combinations of kernel functions 2

1/1 point (graded)
Suppose that k_1 is an isotropic kernel function, but k_2 is a stationary kernel function.

What can we say about the product

$$k\left(Z_1, Z_2\right)=k_1\left(Z_1, Z_2\right) \times k_2\left(Z_1, Z_2\right)$$

It is

- [x] Stationary
- [ ] Isotropic
- [ ] None of the above

✔

Solution:

Although all isotropic kernel functions must be stationary, not all stationary kernel functions are isotropic. Thus, we can only say that a combination of a stationary and isotropic kernel will be stationary.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

For a comprehensive study of covariance functions and kernels, the reader should check Chapter 4 in Williams, Christopher KI, and Carl Edward Rasmussen. Gaussian processes for machine learning. Vol. 2. No. 3. Cambridge, MA: MIT Press, 2006.

Discussion

Hide Discussion

Topic: Module 5: Environmental Data and Gaussian Processes:Spatial Prediction / 5. Effects of Parameters on Kernel Functions

Add a Post

Show all posts by recent activity

- Too difficult compared to video
  I feel like the theoretical question are way too difficult compared to the video courses. 1
- Is the length parameter constant?
  In the definition of RBF kernel, could the length parameter (l) be variable depending on x and x' ? If we apply GP with graph network ... 1

< Previous Next >

Connect

- Blog
- Contact Us
- Help Center
- Media Kit
- Donate

GET IT ON Google play

Download on the App Store