Probabilistic Shape Modelling

Gaussian processes – Deeper insights

Marcel Lüthi

Reminder: Gaussian process

A Gaussian process $p(u) = GP(\mu, k)$

is a probability distribution over functions

$$u : \mathcal{X} \to \mathbb{R}^d$$

such that every finite restriction to function values

$$u_X = \big(u(x_1), \dots, u(x_n)\big)$$

is a multivariate normal distribution

$$p(u_X) = N(\mu_X, k_{XX}).$$

Reminder: Defining a Gaussian process

A Gaussian process

$$GP(\mu, k)$$

is completely specified by a mean function μ and covariance function (or kernel) k.

- $\mu: \mathcal{X} \to \mathbb{R}^d$ defines how the average deformation looks like
- $k: \mathcal{X} \times \mathcal{X} \to \mathbb{R}^{d \times d}$ defines how it can deviate from the mean
  - Must be positive semi-definite

Reminder: Positive definiteness

- A kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^{d \times d}$ is called positive semi-definite, if it gives rise to a positive-semi-definite kernel matrix K_{XX} with

$$K_{ij} = k(x_i, x_j), \qquad i, j = 1, \dots, n$$

for any choice of n and $X = (x_1, \dots, x_n)$

- Exactly what is needed to define a valid covariance matrix!

Reminder: The Karhunen-Loève expansion

We can write $$\mathrm{u} \sim GP(\mu, k)$$

as $$u \sim \mu + \sum_{i=1}^{\infty} \alpha_i \sqrt{\lambda_i}\, \phi_i, \;\; \alpha_i \sim N(0,1)$$

- ϕ_i is the eigenfunction with associated eigenvalue λ_i of the linear operator

$$[T_k u](x) = \int k(x,s)u(s)ds$$

Every sample is a linear combination of Eigenfunctions ϕ_i

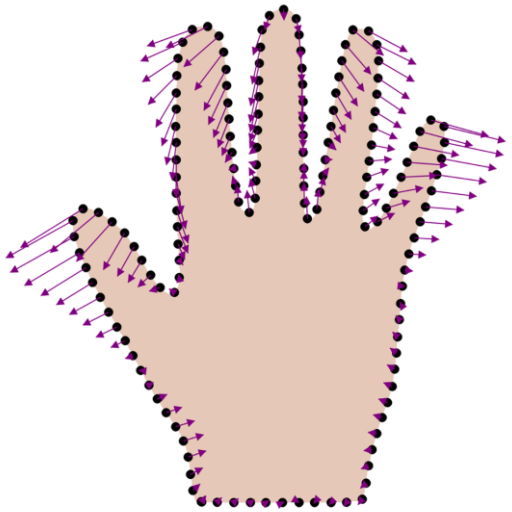
The space of samples

Scalar-valued Gaussian processes

Vector-valued (this course)

- Samples u are deformation fields:

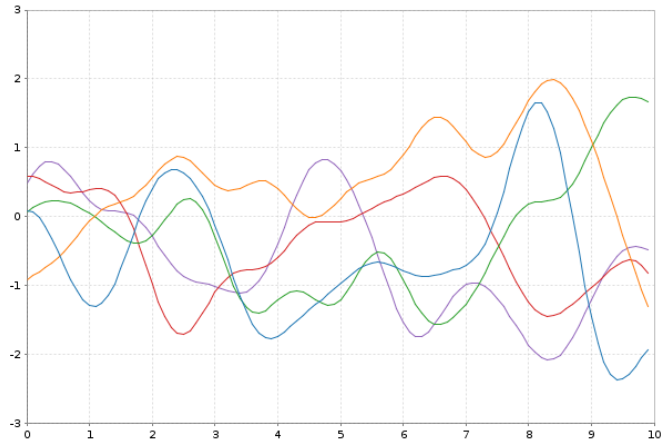
$$u: \mathbb{R}^n \to \mathbb{R}^d$$



Scalar-valued (more common)

- Samples f are real-valued functions

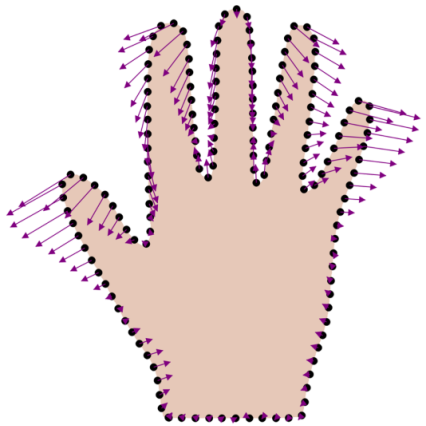
$$f : \mathbb{R}^n \to \mathbb{R}$$



Scalar-valued Gaussian processes

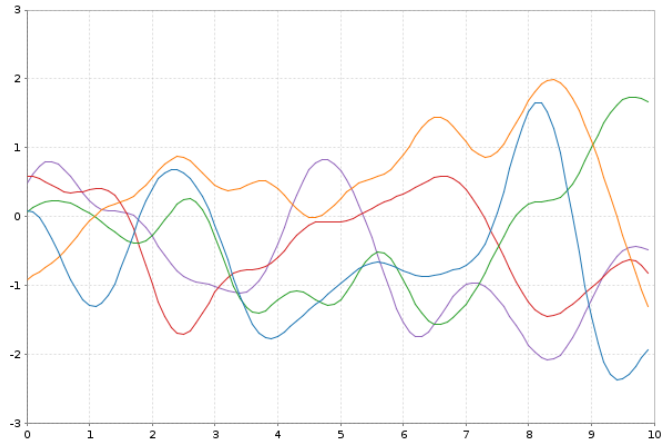
Vector-valued (this course)

$$u \sim GP(\vec{\mu}, \boldsymbol{k})$$
$$\vec{\mu}\colon \mathbb{R}^n \to \mathbb{R}^d$$
$$\boldsymbol{k}\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^{d\times d}$$



Scalar-valued (more common)

$$f \sim GP(\mu, k)$$
$$\vec{\mu}\colon \mathbb{R}^n \to \mathbb{R}$$
$$k\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$



A connection

Matrix-valued kernels can be reinterpreted as scalar-valued kernels:

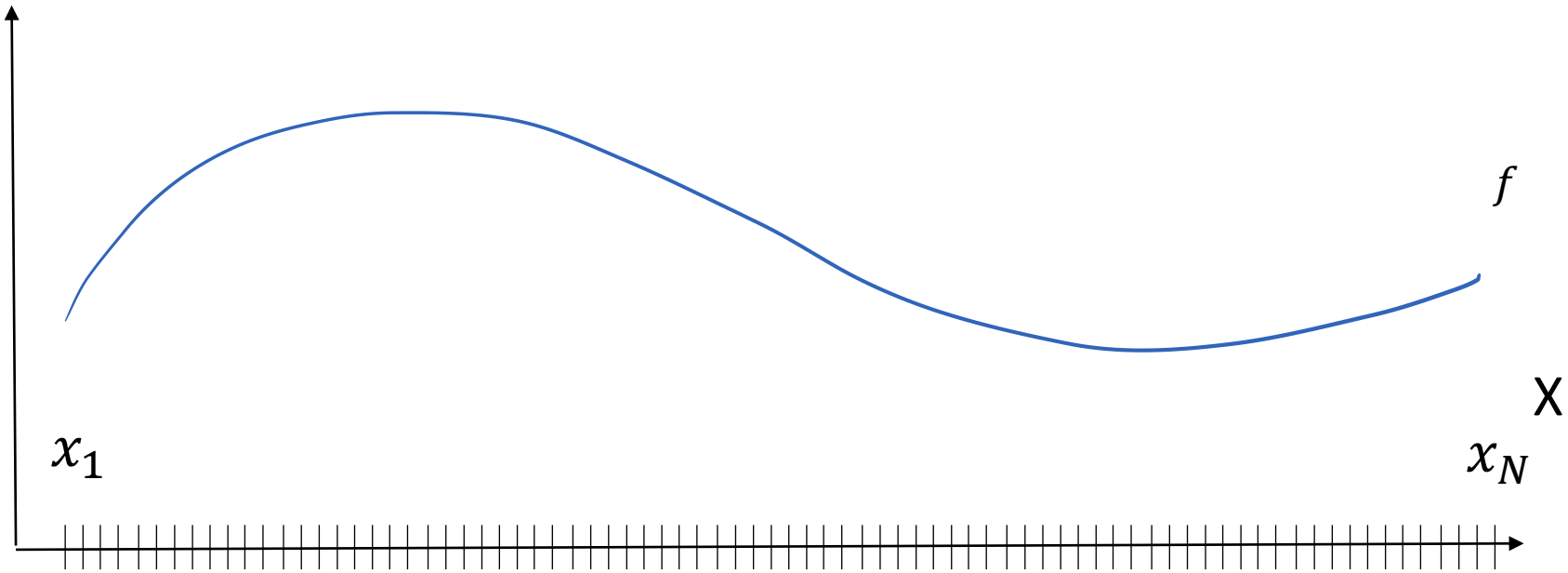
Matrix valued kernel: $\boldsymbol{k}$: $\mathbb{R}^{\boldsymbol{n}} \times \mathbb{R}^{\boldsymbol{n}} \to \mathbb{R}^{\boldsymbol{d}\times\boldsymbol{d}}$

Scalar valued kernel: k: $\mathbb{R}^{\boldsymbol{n}} \times (1..d) \times \mathbb{R}^{\boldsymbol{n}} \times (1..d) \to \mathbb{R}$

Bijection: Define $k((x,i),((x',j)) = \boldsymbol{k}(x',x')_{i,j}$

Everything we know about scalar-valued Gaussian processes can be transferred to the vector-valued case and vice versa.

Simplified mathematical setting – Finite domains



- We consider now a **fixed**, finite domain
- Functions become vectors

Drawing samples

Sampling from $GP(\mu, k)$ defined on a finite domain is done using the corresponding normal distribution $N(\vec{\mu}, \mathrm{K})$

Algorithm for generating random sample $\vec{s}$

1. Do an SVD: $\mathrm{K} = U\Lambda U^T$
2. Draw a normal vector $\vec{\alpha} \sim N(0, I_{n\times n})$
3. Compute $\vec{\mu} + U\Lambda^{\frac{1}{2}}\vec{\alpha}$

The space of samples

$$\mathrm{K} = U\Lambda U^T$$
$$\Rightarrow KU = U\Lambda U^T U = U\Lambda \quad \text{(right multiplication with U orthogonality of U)}$$
$$\Rightarrow KU\Lambda^{-\frac{1}{2}} = U\Lambda^{-\frac{1}{2}} \quad \text{(right multiplication with } \Lambda^{-\frac{1}{2}}\text{)}$$

For a sample $\vec{s}$ we have

$$\vec{s} = \vec{\mu} + U\Lambda^{-\frac{1}{2}}\vec{\alpha} = \vec{\mu} + \mathrm{K}U\Lambda^{-\frac{1}{2}}\vec{\alpha} = \vec{\mu} + \mathrm{K}\vec{\beta}$$

with $with\ \beta = \left(U\Lambda^{-\frac{1}{2}}\alpha\right)$

The space of samples

$$\vec{s} = \vec{\mu} + U\Lambda^{-\frac{1}{2}}\vec{\alpha} = \vec{\mu} + \mathrm{K}\vec{\beta}$$

Writing matrix-multiplication as explicit sum

$$\vec{s} = \vec{\mu} + \sum_i \alpha_i \sqrt{\lambda_i}\, U_{(i,\cdot)}$$

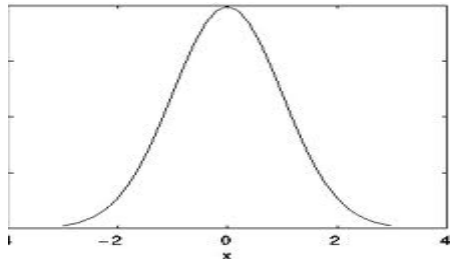
$$\vec{s} = \vec{\mu} + \sum_i K_{(i,\cdot)}\beta_i = \vec{\mu} + \sum_i K_{(\cdot,i)}\beta_i$$

Sample space consists of

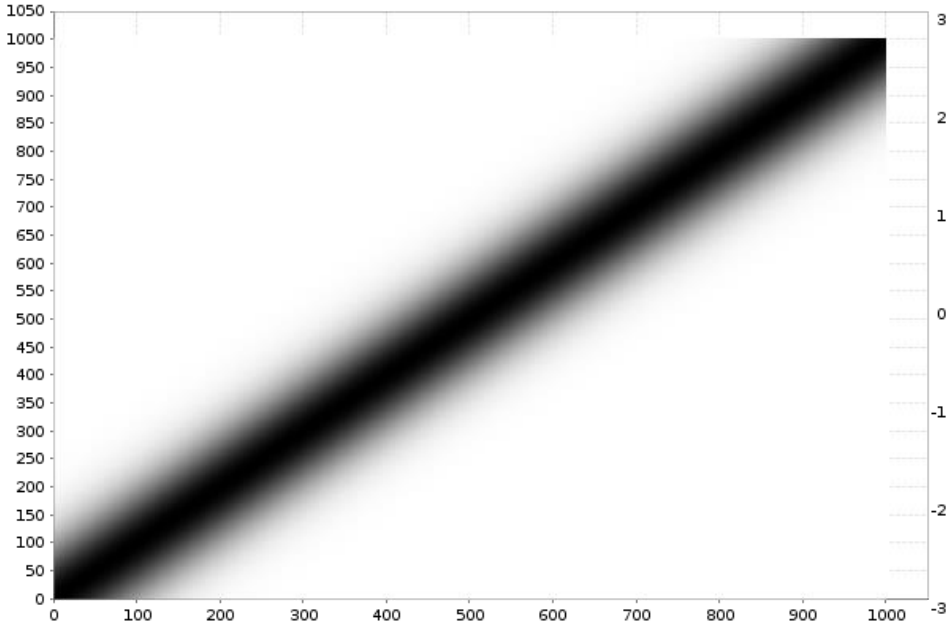
- space of all linear combination of eigenvector of K *(KL-Expansion)*
- space of all linear combinations of columns/rows of K *(RKHS-View)*

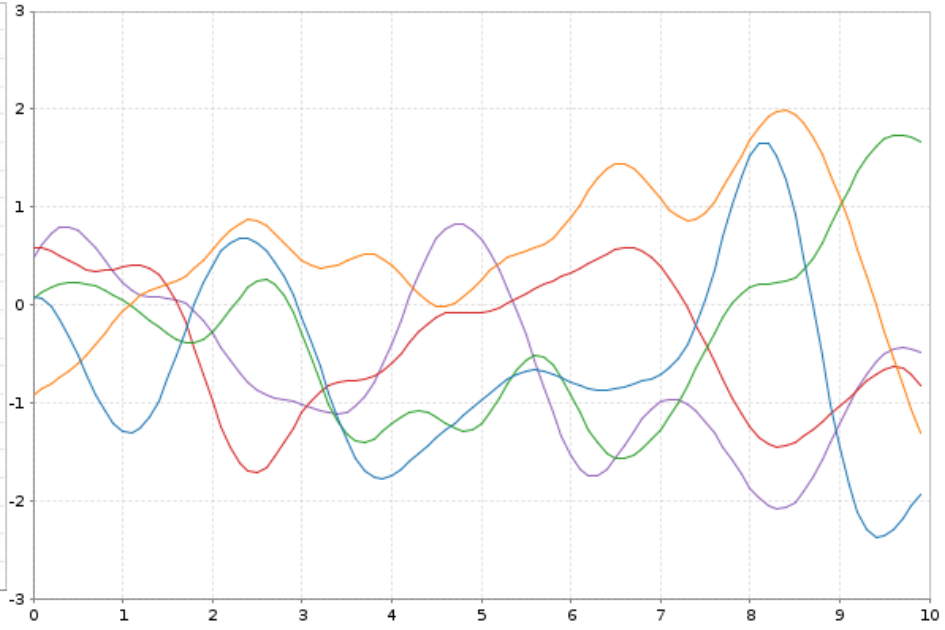
Example: Gaussian kernel

$$k(x, x') = \exp\left(-\frac{\|x - x'\|^2}{\sigma^2}\right)$$



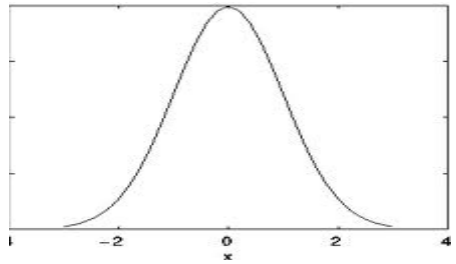
$$\sigma = 1$$



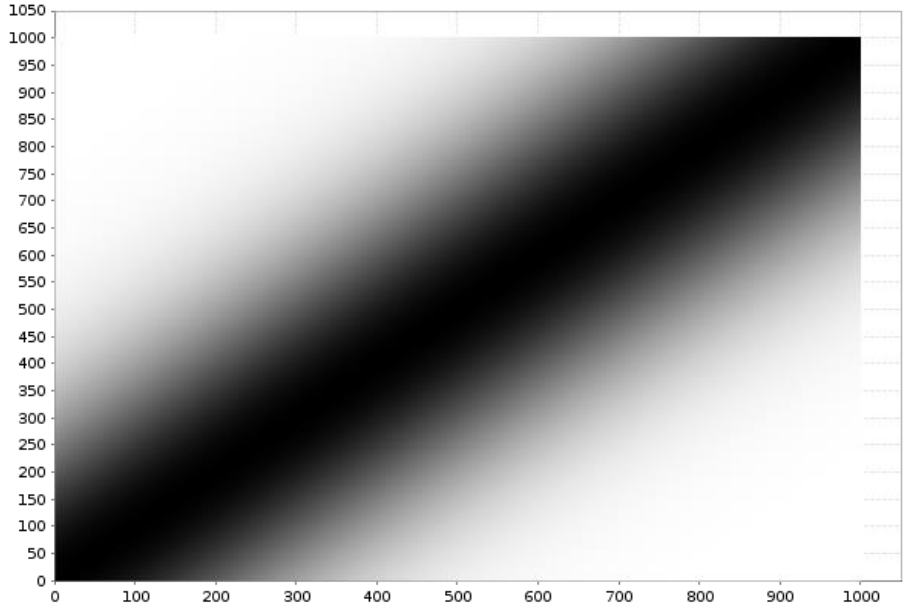


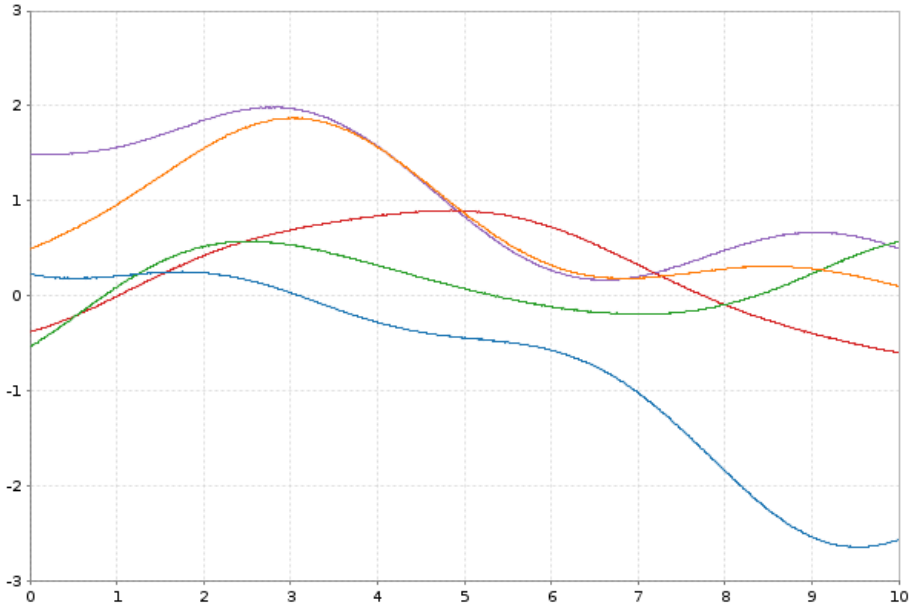
Example: Gaussian kernel

$$k(x, x') = \exp\left(-\frac{\|x - x'\|^2}{\sigma^2}\right)$$



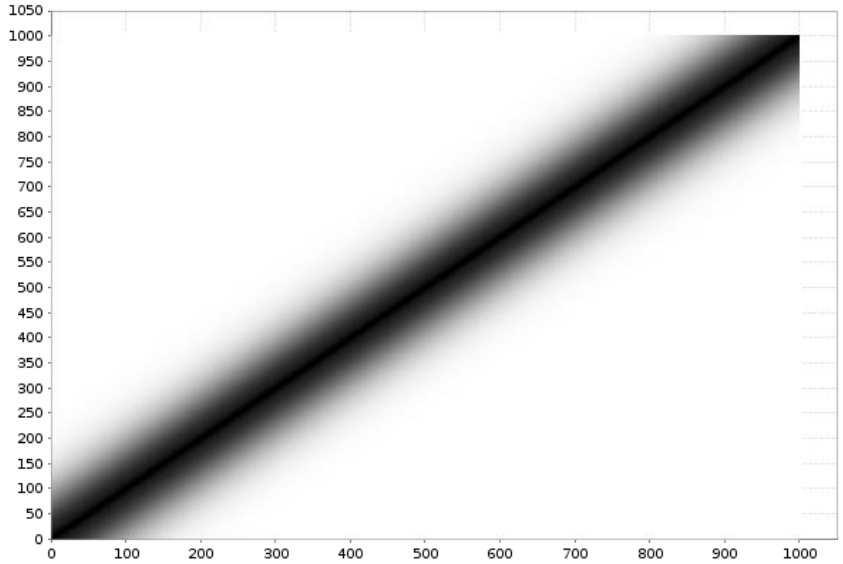
$$\sigma = 3$$

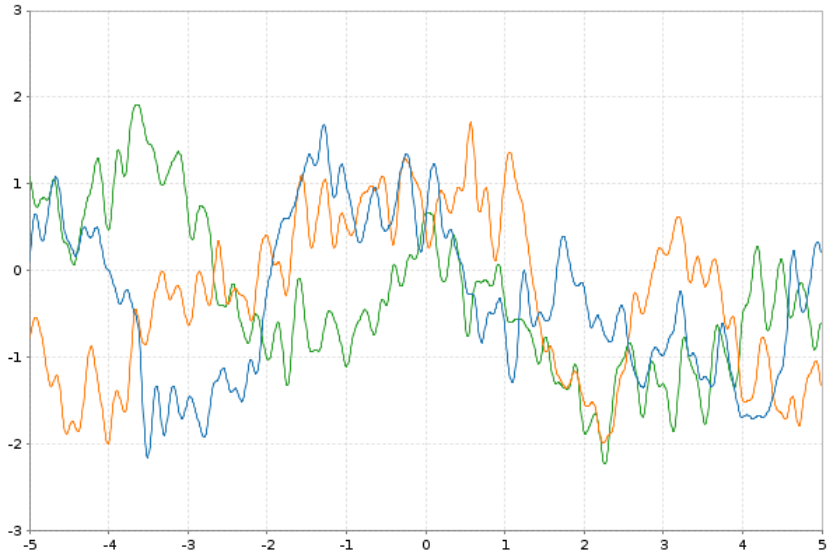




Multi-scale signals

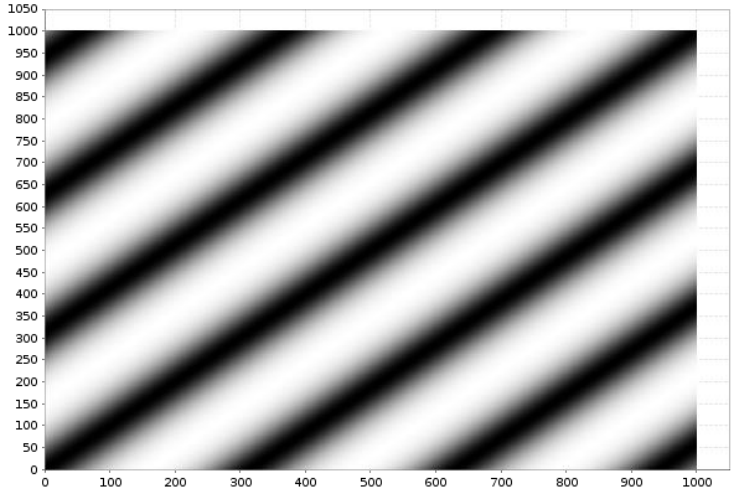
- $\mathrm{k}(\mathrm{x}, \mathrm{x}') = \exp\left(-\left\|x - \frac{x'}{1}\right\|^2\right) + 0.1 \exp\left(-\left\|x - \frac{x'}{0.1}\right\|^2\right)$

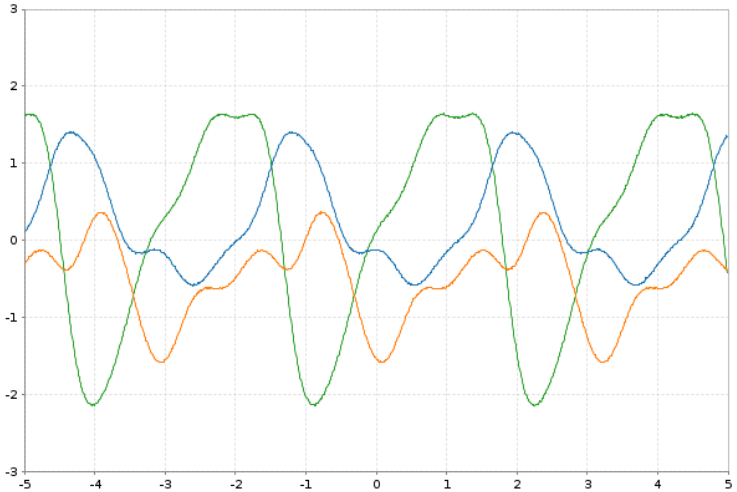




Periodic kernels

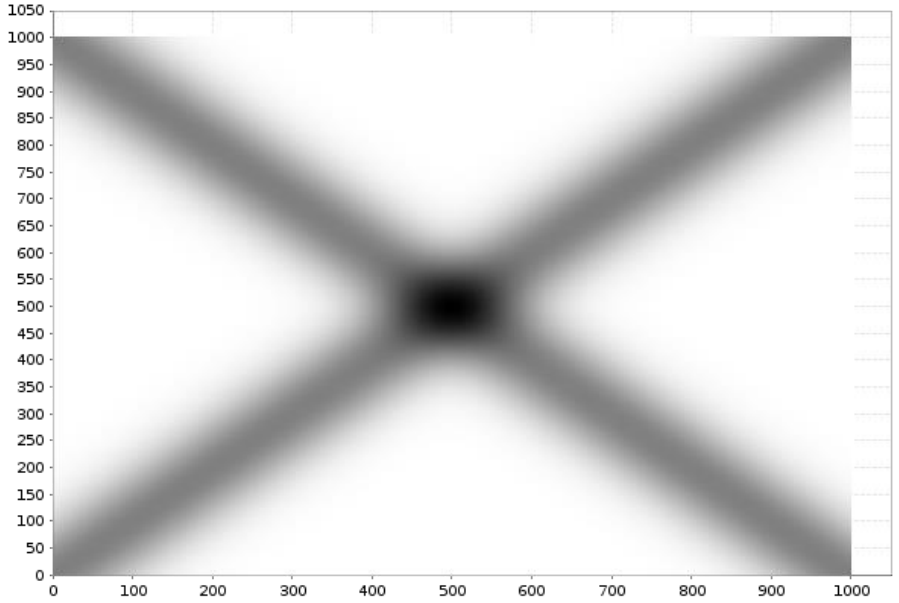
- Define $u(x) = \begin{pmatrix} \cos(x) \\ \sin(x) \end{pmatrix}$
- $k(x, x') = \exp(-\|(u(x) - u(x')\|^2 = \exp(-4 \sin^2\left(\frac{\|x - x'\|}{\sigma^2}\right))$

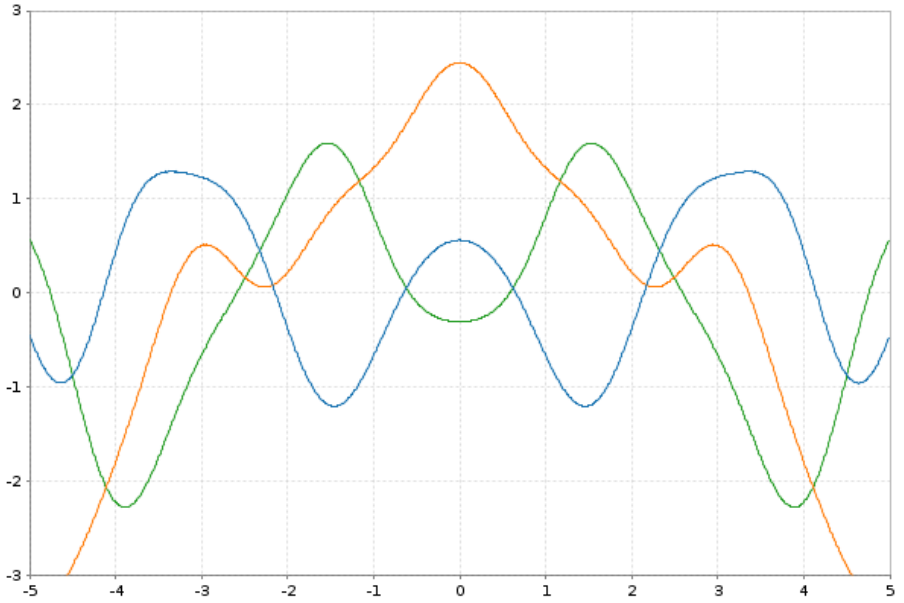




Symmetric kernels

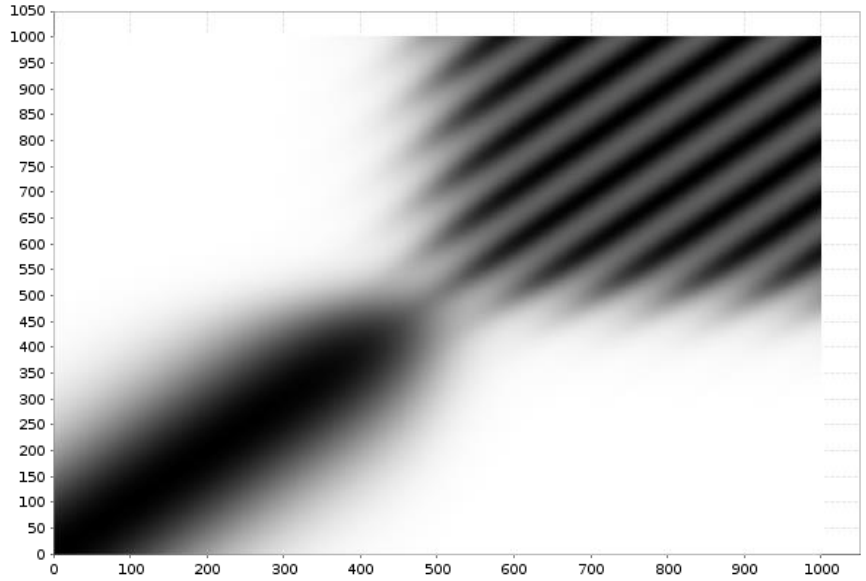
- *Enforce that f(x) = f(-x)*
- $k(x, x') = k(-x, x') + k(x, x')$

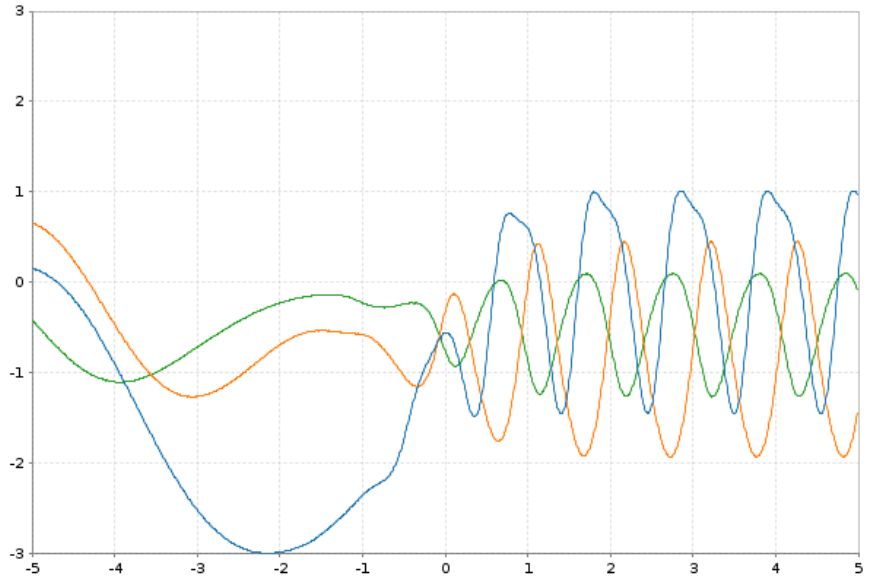




Changepoint kernels

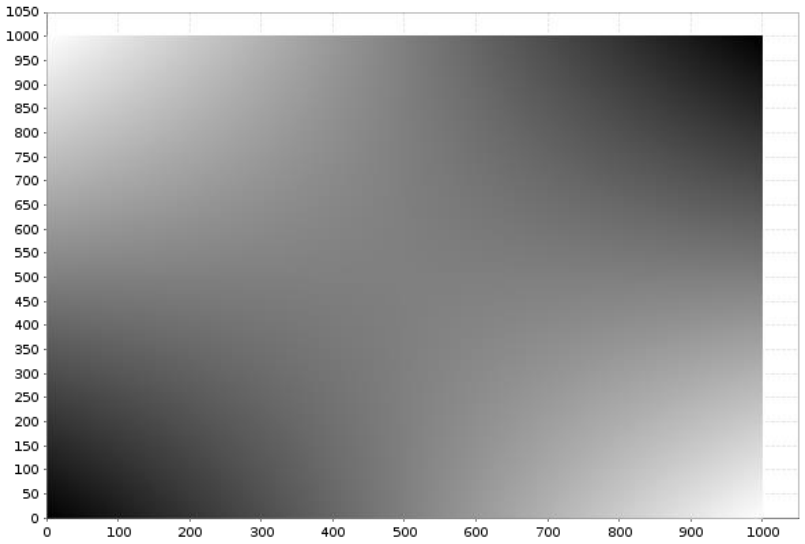
- $k(x, x') = s(x)k_1(x, x')s(x') + (1 - s(x))k_2(x, x')(1 - s(x'))$
- $\mathrm{s}(x) = \frac{1}{1+\exp(-x)}$

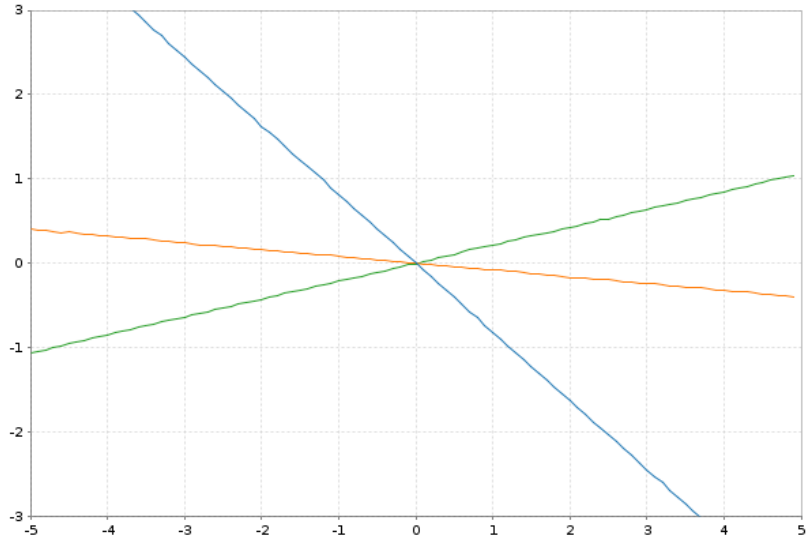




Combining existing functions

$$k(x, x') = f(x)f(x')$$

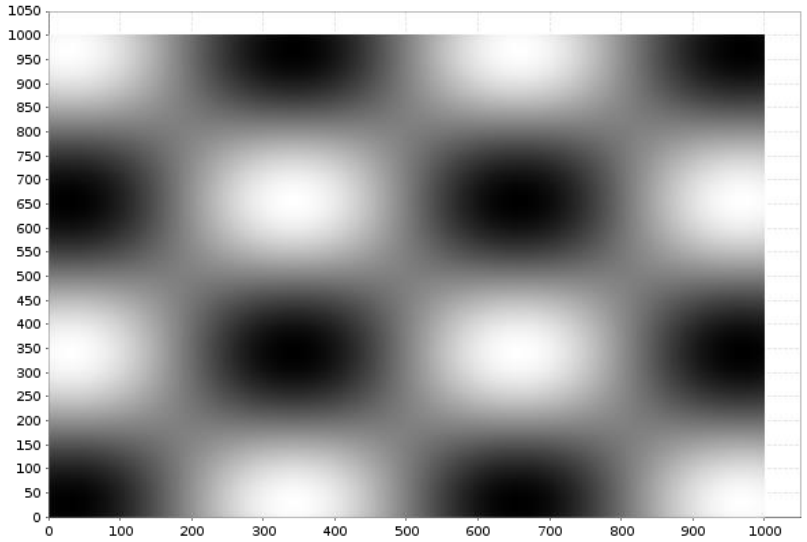


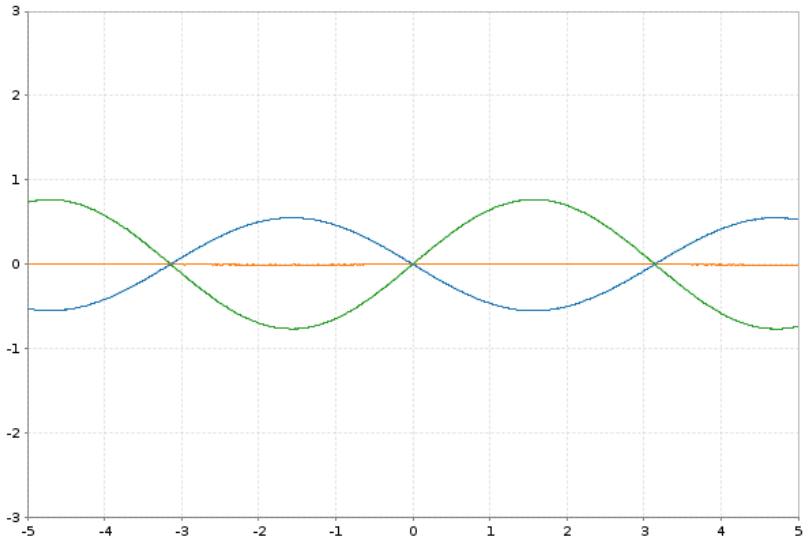


$$f(x) = x$$

Combining existing functions

$$k(x, x') = f(x)f(x')$$

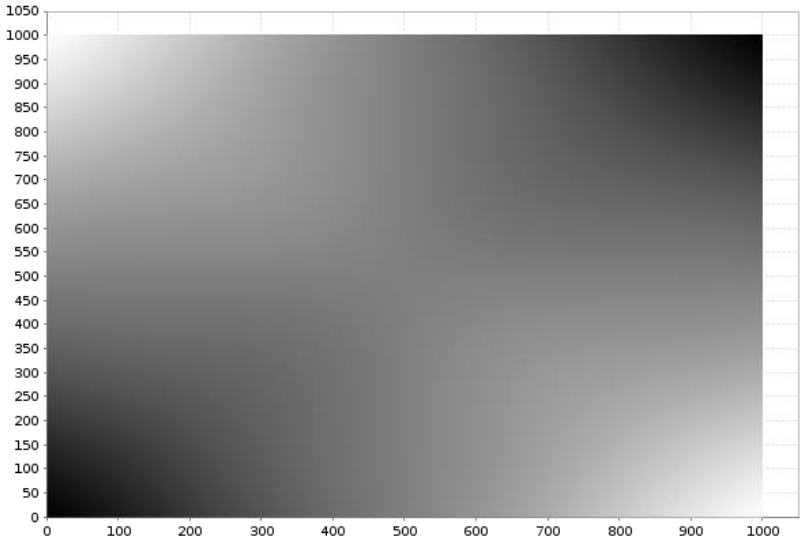


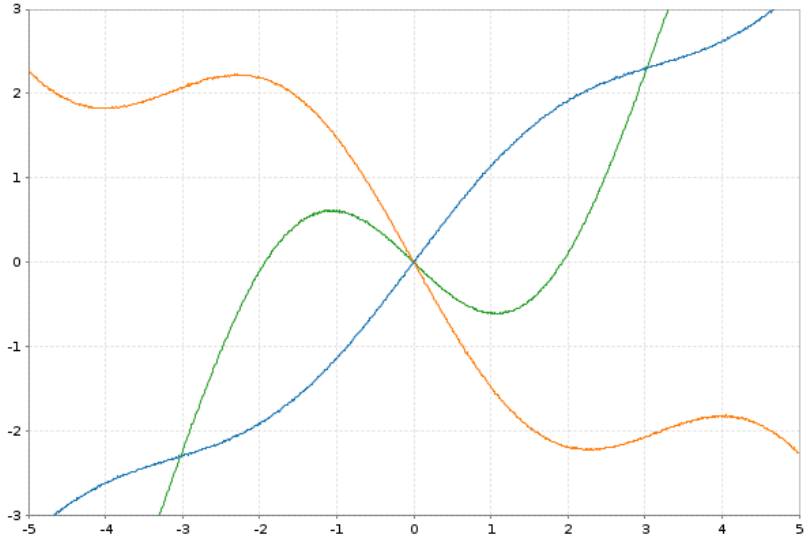


$$f(x) = \sin(x)$$

Combining existing functions

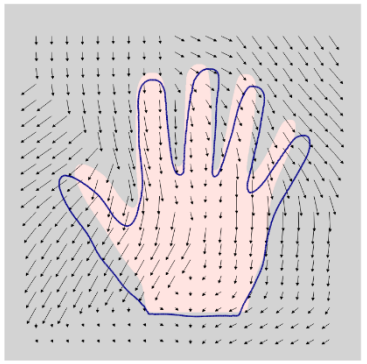
$$k(x, x') = \sum_i f_i(x) f_i(x')$$



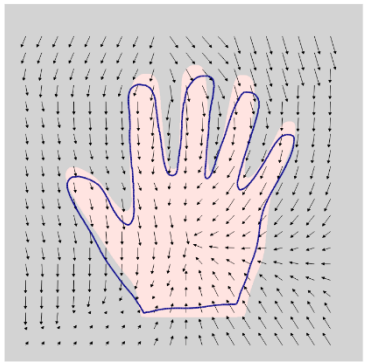


$$\{f_1(x) = x, f_2(x) = \sin(x)\}$$

Statistical models

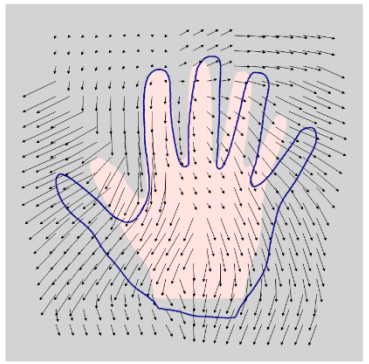


$u^1 : \Omega \to \mathbb{R}^2$



$u^2 : \Omega \to \mathbb{R}^2$

...



$u^n : \Omega \to \mathbb{R}^2$

$$\mu(x) = \overline{u}(x) = \frac{1}{n}\sum_{i-1}^{n} u^i(x)$$

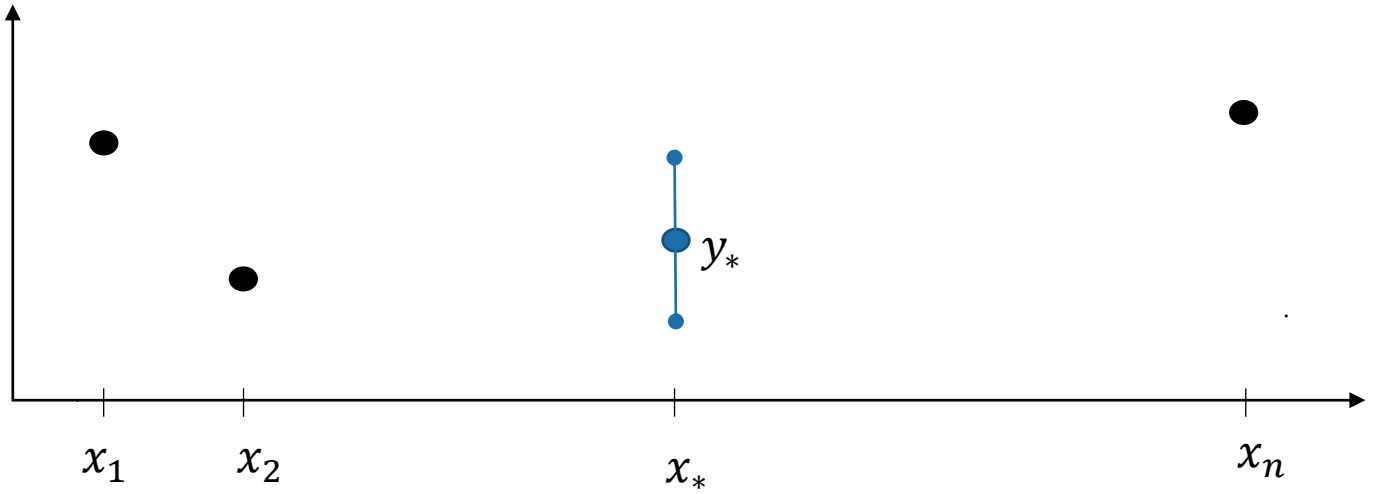
$$k_{SM}(x, x') = \frac{1}{n-1}\sum_{i}^{n} (u^i(x) - \overline{u}(x))\left(u^i(x') - \overline{u}(x')\right)^T$$

Statistical shape models are linear combinations of example deformations $u^1, \dots u^n$.

Gaussian process regression

Gaussian process regression

- Given: observations $\{(x_1, y_1), \dots, (x_n, y_n)\}$
- Model: $y_i = f(x_i) + \epsilon,\ f \sim GP(\mu, k)$
- Goal: compute $p(y_* | x_*, x_1, \dots, x_n, y_1, \dots, y_n)$



Gaussian process regression

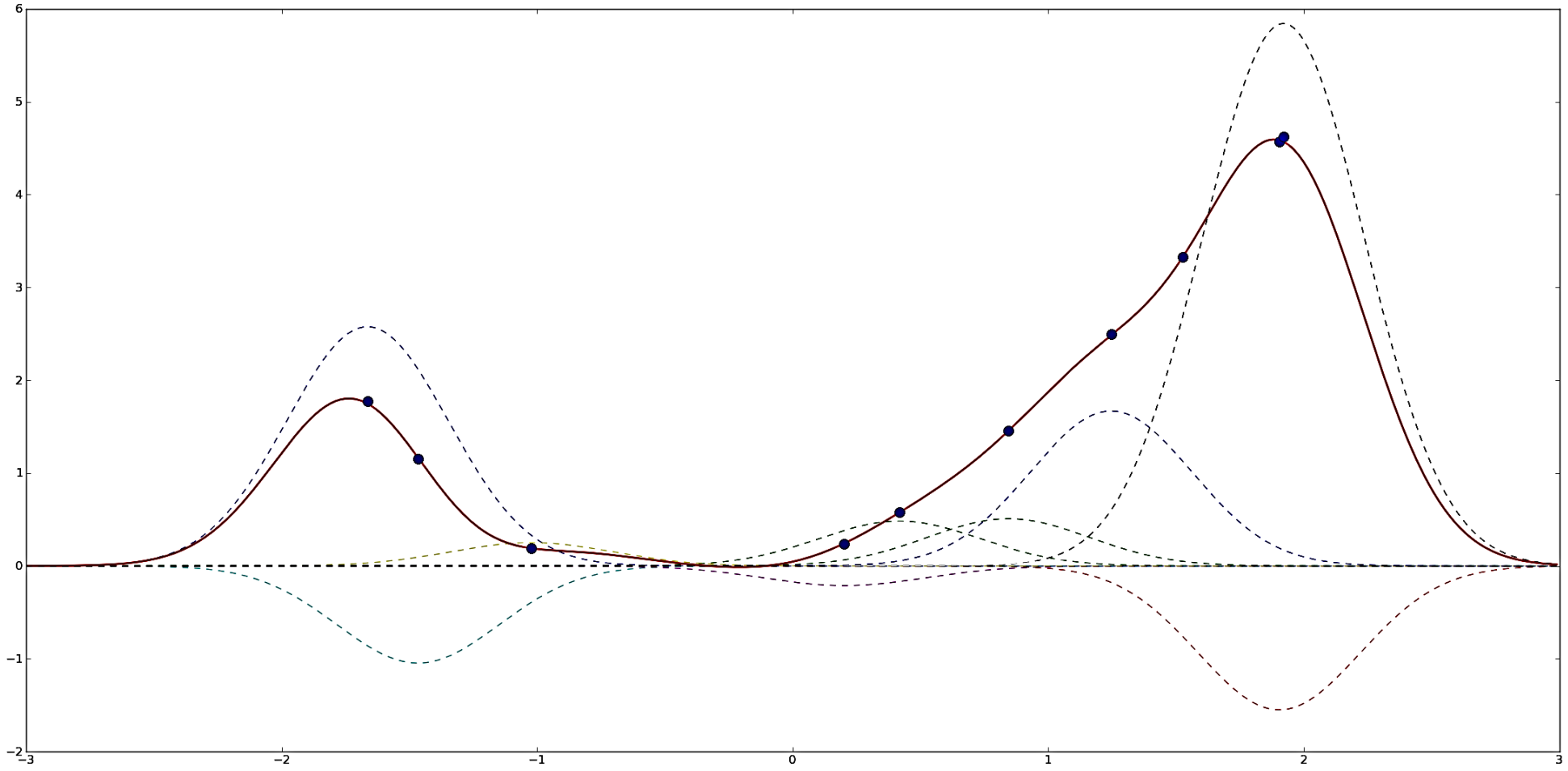
- Solution given by posterior process $GP(\mu_p, k_p)$ with

$$\mu_p(x_*) = K(x_*, X)[K(X, X) + \sigma^2 I]^{-1} y$$

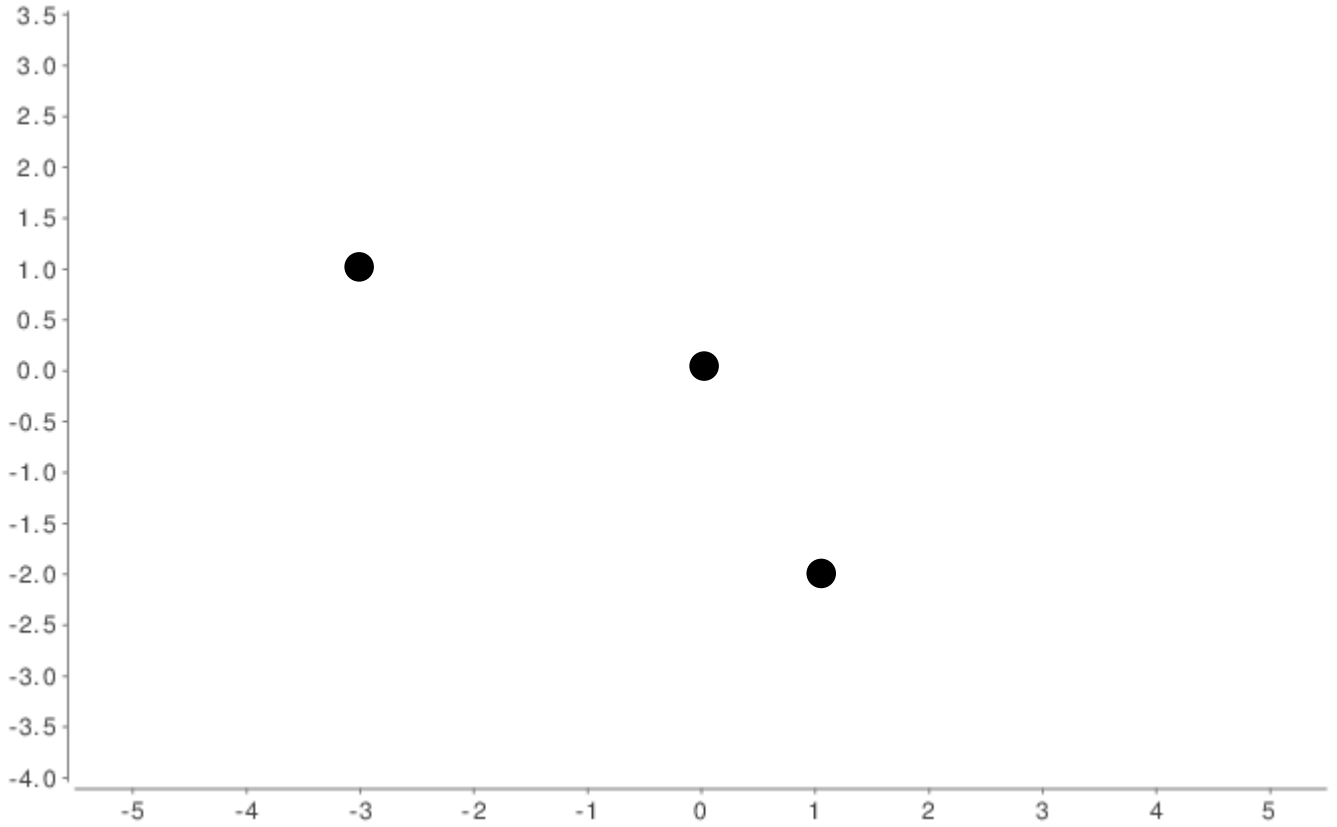
$$k_p(x_*, x_*') = k(x_*, x_*') - K(x_*, X)[K(X, X) + \sigma^2 I]^{-1} K(X, x_*')$$

- The covariance is independent of the value at the training points
  - Structure of posterior GP determined solely by kernel.
- The most likely solution is a linear combination of kernels evaluated at the training points
  - This is known as the **Representer Theorem** in machine learning.
  - Structure of solution determined solely by kernel.

Illustration: Representer theorem

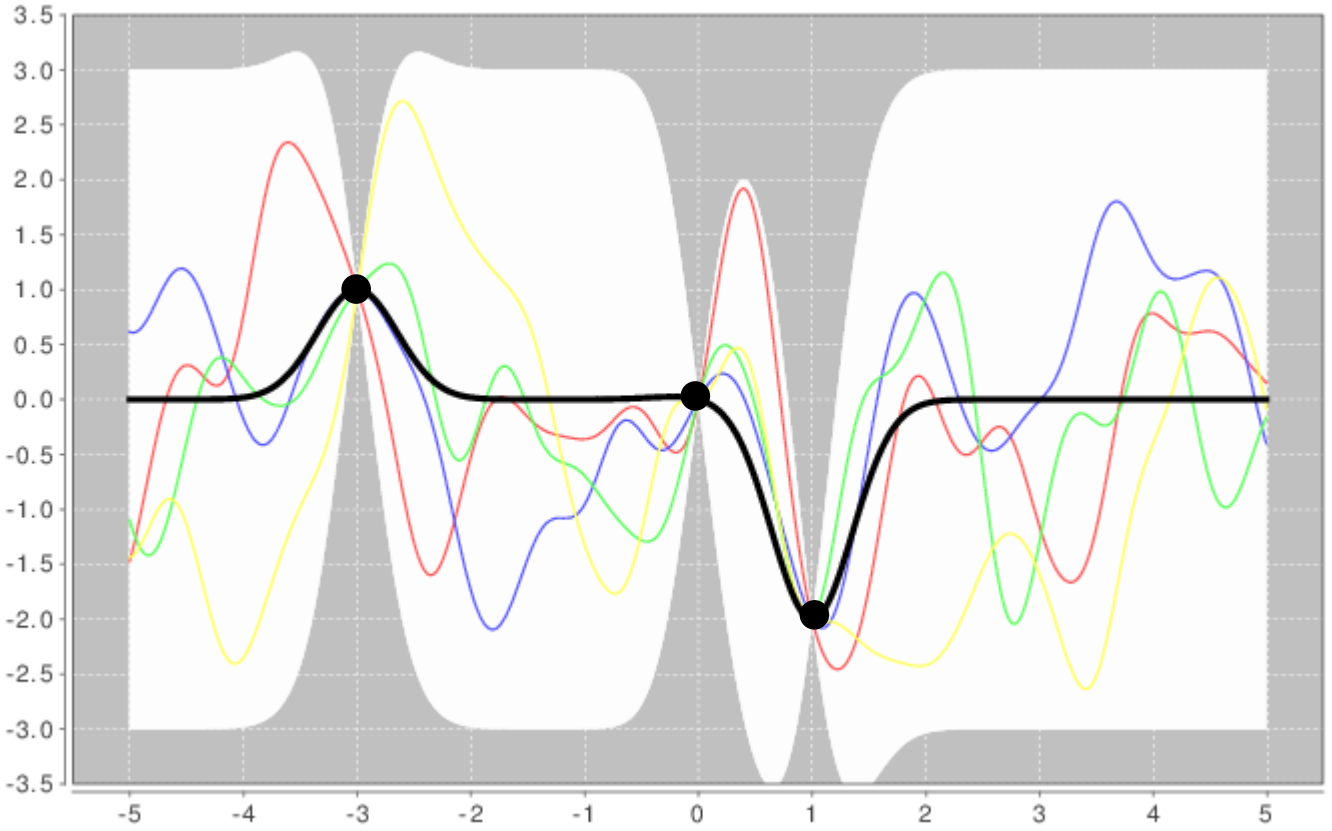


Examples



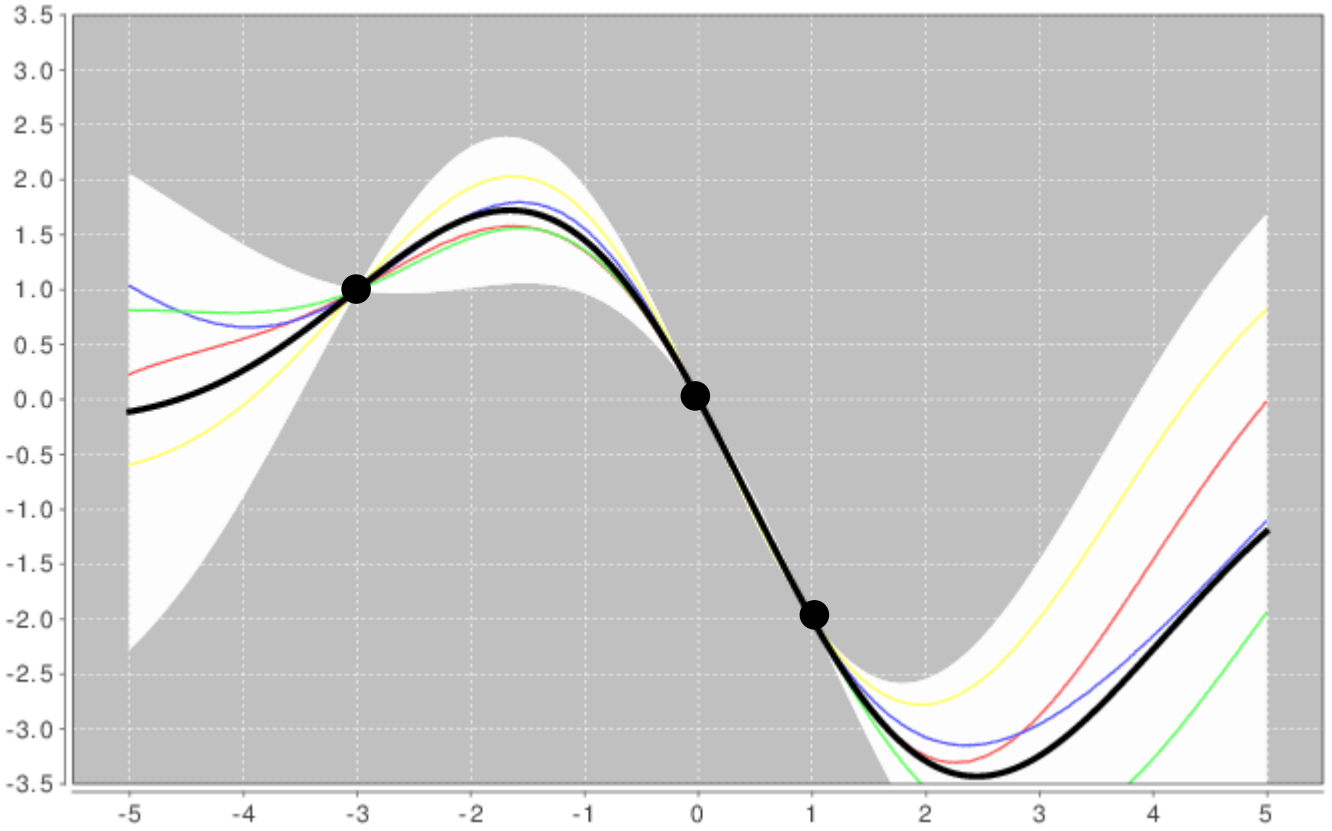
Examples

- Gaussian kernel ($\sigma = 1$)



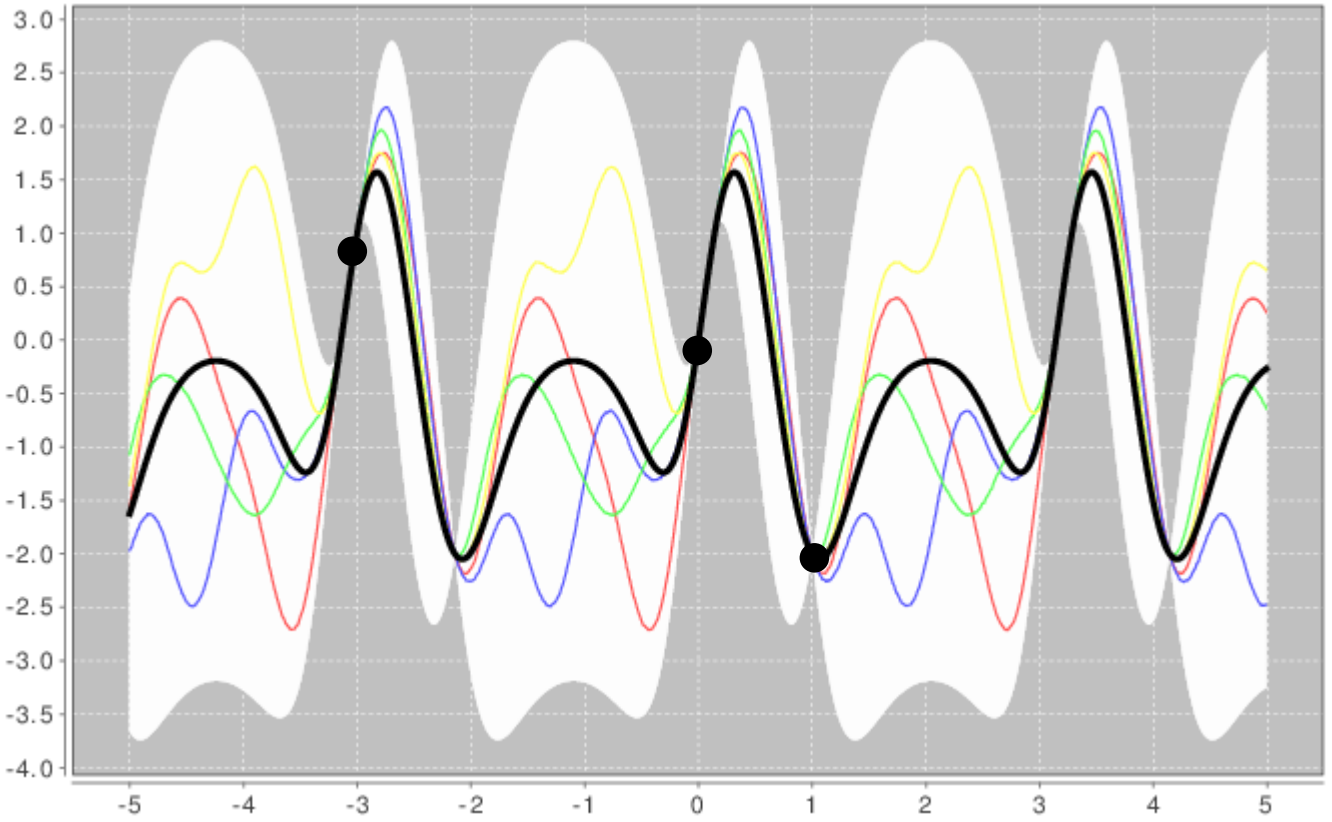
Examples

- Gaussian kernel ($\sigma = 5$)



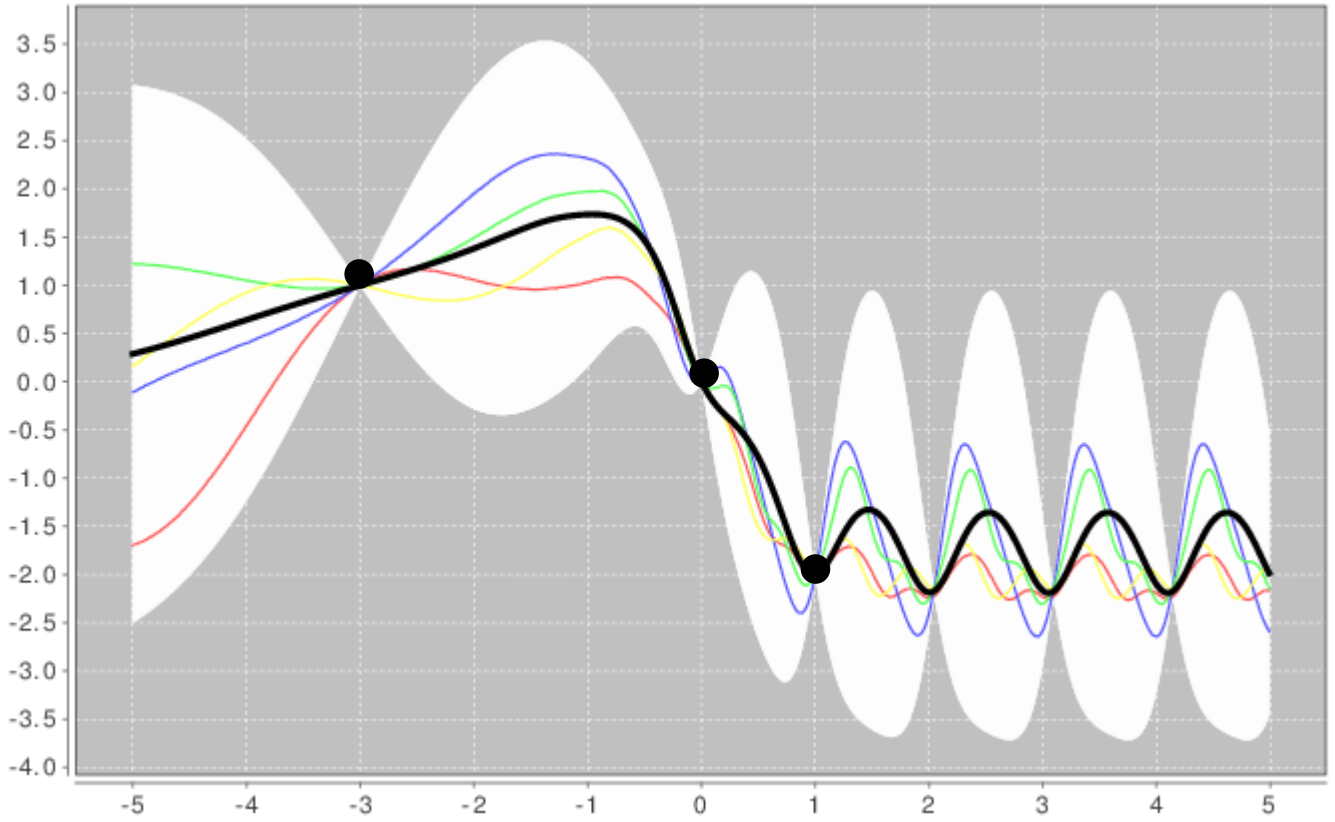
Examples

- Periodic kernel



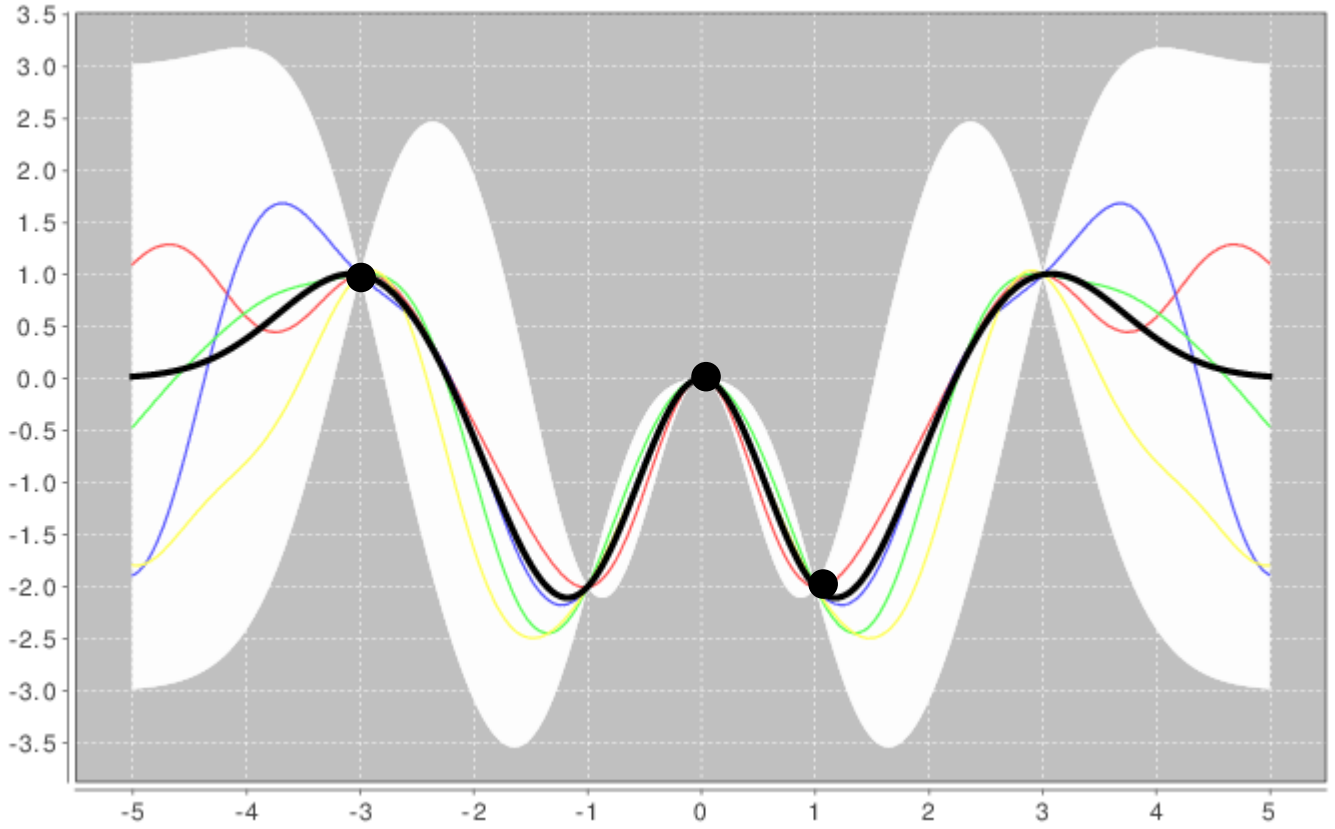
Examples

- Changepoint kernel



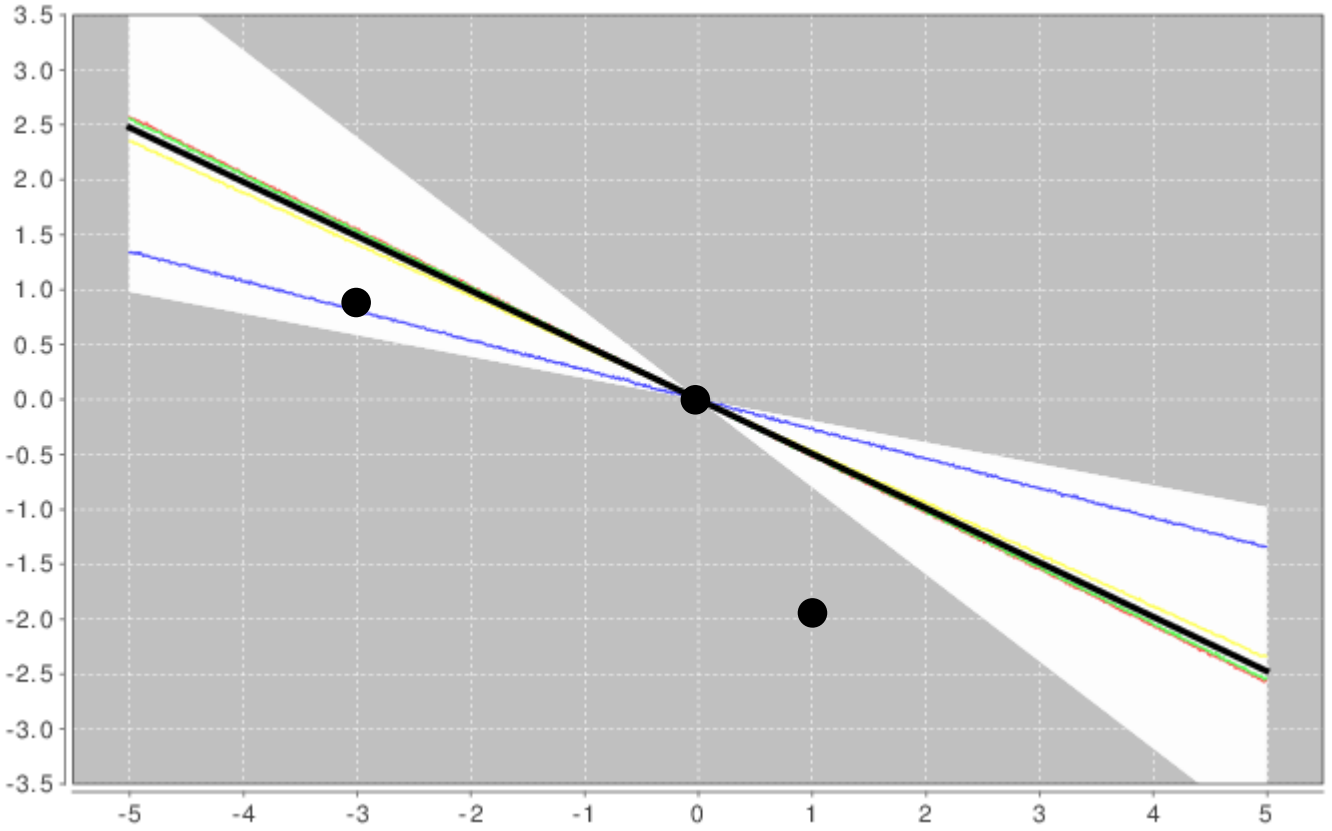
Examples

- Symmetric kernel



Examples

- Linear kernel



Summary – Gaussian processes

Sample Space of a GP. Two views

1. KL-Expansion:
   - Global functions that best capture the global properties of the GP

2. Linear combinations of the kernels $k(\cdot, x)$, fixed at point x
   - Apply the kernels locally at each point
   - Properties of kernels - Regularity/smoothness directly transferred to samples

- In inference tasks, the structure of the kernel determines the prediction
  - => Extremely important to model it well