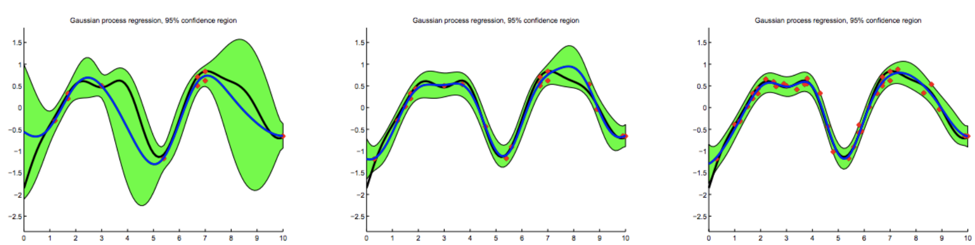
gaussian processes

- Lets say a function maps $x_1 ... x_m$ onto the set of real numbers $\mathcal{R}$. probability distributions over functions with finite domains such as the one above can be represented using a finite-dimensional multivariate Gaussian distribution over function outputs $f(x_1), ..., f(x_m)$ at a finite number of input points $x_1, x_2 ... x_m$. Each vector corresponds to one such function and gets assigned a probability density using the formula for pdf in multivariate gaussians.
- for probability distributions over functions with infinite domains (unlike $x_1 ... x_m$) we turn to something called gaussian processes.
- In particular, a collection of random variables {f(x) : x ∈ X } is said to be drawn from a Gaussian process with mean function m(·) and covariance function k(·, ·) if for any finite set of elements $x_1 ... x_m$ ∈ X , the associated finite set of random variables $f(x_1), ..., f(x_m)$ have distribution:

$$\begin{bmatrix} f(x_1) \\ \vdots \\ f(x_m) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(x_1) \\ \vdots \\ m(x_m) \end{bmatrix}, \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_m) \\ \vdots & \ddots & \vdots \\ k(x_m, x_1) & \cdots & k(x_m, x_m) \end{bmatrix} \right).$$

- Observe that the mean function and covariance function are aptly named since the above properties imply that:
  - $m(x) = E[x]$
  - $k(x, x') = E[(x - m(x))(x' - m(x'))]$
- Look at the following visualization of gaussian processes. While interpreting this visualization keep in mind that a gaussian process is a probability distribution over infinite dimensional functions.



(a) (b) (c)

Figure 3: Gaussian process regression using a zero-mean Gaussian process prior with $k_{SE}(\cdot, \cdot)$ covariance function (where $\tau = 0.1$), with noise level $\sigma = 1$, and (a) $m = 10$, (b) $m = 20$, and (c) $m = 40$ training examples. The blue line denotes the mean of the posterior predictive distribution, and the green shaded region denotes the 95% confidence region based on the model's variance estimates. As the number of training examples increases, the size of the confidence region shrinks to reflect the diminishing uncertainty in the model estimates. Note also that in panel (a), the 95% confidence region shrinks near training points but is much larger far away from training points, as one would expect.

- The green region above indicates a confidence interval within which the function lies.