

Gaussian Process Regression

As described in the previous section, we make use of a surrogate model to give an intermediate view of the energy landscape. We construct this surrogate, using a method known as *gaussian process regression* (GPR). This method has the benefits of being parameter free, simple and the ability to interpolate data perfectly. In this section I will provide a brief overview of the method as well as some examples. We start with some definitions,

1.0.1 Gaussian Process

A gaussian process is a special case of so-called random processes. A random process is an infinite collection of random variables $\{X_t, t \in T\}$, where T is some (continuous) index set. The way in which the X_T relate to one another, determines the type of random process. For gaussian processes the joint distribution of any finite subset $X_{1,\dots,n}$ is a multivariate normal gaussian,

$$X_{1,\dots,n} = (X_1, \dots, X_n) \sim N(0, \Sigma).$$

Where Σ is the covariance matrix, which can be calculated using the GP's associated *kernel* K in the following way,

$$\Sigma_{ij} = K(X_i, X_j).$$

Where the kernel is some positive function. The behaviour of a GP, is therefore completely determined by the choice of kernel. A GP, can be viewed as a distribution over functions, where the functions domain is the index set T and its image, the possible realizations of the variables X_t . The general shape of the function, whether it is continuous, periodic and so forth, is determined by the kernel. A common choice of kernel, is the *radial basis function*

$$K_{RBF}(X_i, X_j) = \exp\left(-\frac{|X_i - X_j|^2}{\theta}\right),$$

which has the nice property of being infinitely differentiable. The constant θ is a hyperparameter, which is used to tune the kernel. A variety of sampled functions is shown in the FIG.

1.0.2 Regression

In general, we are not interested in sampling random functions. Instead we want to use our GP to make predictions. Let us assume that we want to model a real-valued function f , whose domain is an interval $I \subset \mathbb{R}$. Let $X = \{X_1, \dots, X_n\}$ be our data points and $f(X)$ their "true" values. Our goal is to use this data to predict the value of f at a new point \tilde{x} . The first step, is to construct the joint probability distribution,

$$(X, \tilde{x}) \sim N\left(0, \begin{bmatrix} \Sigma_{XX} & \Sigma_{X\tilde{x}} \\ \Sigma_{\tilde{x}X} & \Sigma_{\tilde{x}\tilde{x}} \end{bmatrix}\right).$$

We can now condition on the true values of X , to obtain a conditional probability distribution for \tilde{x} . I have chosen to leave out the derivation of this distribution as it is a bit cumbersome.

$$(X, \tilde{x}) \mid X = f(X) \sim N(\mu(\tilde{x}), \sigma^2(\tilde{x})).$$

Sampled functions from $\mathcal{GP}(0, \Sigma)$

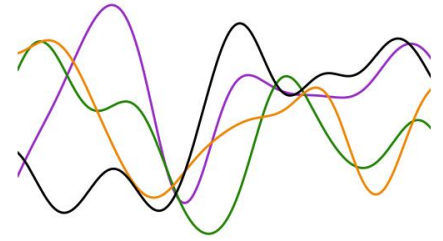


figure: 4 randomly sampled functions from a gaussian process with the radial basis kernel. Since this kernel is differentiable, the sampled functions are smooth.

Gaussian Process Regression

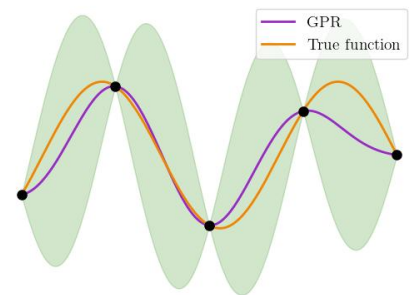


figure: A simple regression for a sine wave (shown in yellow). Notice that the fit (purple) interpolates the data! The uncertainty is displayed in green. As expected, the uncertainty tends to zero in proximity of the data points.

where the mean and variance are given by,

$$\begin{aligned}\mu(\tilde{x}) &= (\Sigma_{\tilde{x}X} \Sigma_{XX}^{-1}) \cdot f(X) \\ \sigma^2(\tilde{x}) &= \Sigma_{\tilde{x}\tilde{x}} - \Sigma_{\tilde{x}X} \Sigma_{XX}^{-1} \Sigma_{X\tilde{x}}.\end{aligned}$$

We have thus obtained a prediction and uncertainty for $f(\tilde{x})$ given our data. FIG illustrates an example. An important part of GPR, is the choice of kernel and hyperparameters. Since these fully determine the functions that a GP is able to reproduce, it is essential that they are chosen wisely. There exist a variety of methods that accomplish this, which I shall not delve into.