Introduction

Input is denoted as x, output as y. The problem at hand is inductive, predictions are made on previously unseen data. Two approaches to supervised learning: restrict the class of functions to consider (i.e. linear functions), or, give a prior probability to a set of functions one considers according to belief in suitability. The problem with the first approach is picking the right set of functions. In the second approach, one can't know which functions are more suitable then others, and one solution is to observe possible functions as a *Gaussian distribution*. A stochastic process governs over functions, as does a random variable over vectors or scalars.

1.1 A Pictorial Introduction to Bayesian Modelling

We start with a 1-d regression problem. Observe figure 1.1. Without any knowledge of the data, one suppouses all functions are equally possible and the expected mean is 0, meaning that the mean of all possible functions randomly selected so far is 0. The shaded region denotes two standard deviations. Now, a dataset of two datapoints is provided, and thus only functions that pass through those datapoints are **only** considered (important to note that this should usually just increase preference around these points). Combining the prior and data creates the *posterior*.

One way to do inference with gaussian processes is to randomly draw functions and inspect how well they fit the data. This is inefficient.

When specifying the prior, it is important for it to have properties suitable for inference. Finding the right functions boils down to getting the properties of the covariance function of the Gaussian process.

When doing (binary) classification, one is looking for the probability that an example input belongs to a class in the interval of $[0, 1]$. This will be done by squashing the ouput to that range using some function (like the logistic function). Now we look at figure 1.2 where a 2d prior has been selected, squashed with the logistic function. A dataset is shown in c), and it's classification confidence output in d).