Small Exercises 5

Linear Regression and Kernels

These exercises are meant as preparation for the inverted classroom lecture. Keep your answers short: two or three sentences, sometimes even less, should suffice.

1 Linear Regression

Problem 1: Let's assume you are using a maximum likelihood version of the regression algorithm with a polynomial basis function. What can happen when the order of your polynomial is considerably higher than the number of training points?

Overfitting

Problem 2: Why is the bayesian formulation useful?

It helps us to explicitely say what the assumptions we are making are. Or it allows us to do online learning much easier. Or it allows us to put priors on our data distributions which is very helpful when we have limited amounts of data. Etc.

Problem 3: Which two assumptions do we make regarding the weight parameters and the data points when we use least squares regression with L_2 regularization?

We assume that the data has a gaussian noise with the mean equal to our predictive function and that the weight parameters have isotropic gaussian noise where the gaussian has zero mean.

2 Kernels

Problem 4: Why are kernels useful? I.e. how do they help us combat the curse of dimensionality?

The dual formulation allows us to be limited by the number of training points, not the number of basis functions.

Problem 5: What is difficult about proving a function is a kernel assuming that you can't use the kernel construction rules in the slide "Making Kernels with Kernels"?

It can be very hard to prove a matrix is positive semi-definite.