16.5 Laplacian Processes

All the prior distributions considered so far are *data independent* priors; in other words, $p(f)$ does not depend on X at all. This may not always be the most desirable choice, thus we now consider *data dependent* priors distributions, $p(f|X)$. This goes slightly beyond the commonly used concepts in Bayesian estimation.

Before we go into the technical details, let us give some motivation as to why the complexity of an estimate can depend on the locations where data occurs, since we are effectively updating our prior assumptions about f after observing the data placement. Note that we do not modify our prior assumptions based on the targets y_i, but rather as a result of the distribution of patterns x_i: Different input distribution densities might for instance correspond to different assumptions regarding the smoothness of the function class to be estimated. For example, it might be be advisable to favor smooth functions in areas where data are scarce, and allow more complicated functions where observations abound. We might not care about smoothness at all in regions where there is little or no chance of patterns occurring: In the problem of handwritten digit recognition, we do not (and should not) care about the behavior of the estimator on inputs x looking like faces.

Finally, we might assume a specific distribution of the *coefficients* of a function via a data-dependent function expansion; in other words, an expansion of f into the span of $\Phi := \{\phi_1, \ldots, \phi_M\}$, where ϕ_i are functions of the observed data X and of x. We focus henceforth on the case where $M = m$ and $\phi_i(x) := k(x_i, x)$.

The specific benefit of this strategy is that it provides us with a correspondence between linear programming regularization (Section 4.9.2) and weight decay regularizers (Section 4.9.1), and Bayesian priors over function spaces, by analogy to regularization in Reproducing Kernel Hilbert Spaces and Gaussian Processes.[10]

16.5.1 Data Dependent Priors

Recall the reasoning of Section 16.1.3. We obtained (16.11) under the assumption that X and f are independent random variables. In the following, we repeat the derivation without this restriction, and obtain

$$p(Y|f, X)p(f|X) = p(Y, f|X), \tag{16.87}$$

and likewise,

$$p(f|Y, X)p(Y|X) = p(Y, f|X). \tag{16.88}$$

Combining these two equations provides us with a modified version of Bayes' rule, which after solving for $p(f|Y, X)$, reads

$$p(Y|f, X)p(f|X) = p(f|X, Y)p(Y|X), \tag{16.89}$$

10. We thank Carl Edward Rasmussen for discussions and suggestions.