1 Gaussian Process

[1] introduces Gaussian Process regression as generalization of Bayesian regression. In linear regression setup, we assume output value $y \in \mathbb{R}$ is a linear function of inputs $x \in \mathbb{R}^d$, corrupted by iid normal noise.

$$y = f(x) + \epsilon \quad \text{where} \quad f(x) = w^T x \quad \text{and} \quad \epsilon \sim \mathcal{N}(0, \sigma_n^2) \tag{1}$$

The Bayesian setup considers $w \in \mathbb{R}^d$ as a random variable, endowed with prior $w \sim \mathcal{N}(0, \Sigma_p)$. Using Bayes rule, we can find the posterior of weights given data, which is again a normal random variable $p(w \mid X, y) = \mathcal{N}\left(w \,; A^{-1}b, A^{-1}\right)$ where $A = \frac{1}{\sigma_n^2} X^T X + \Sigma_p^{-1}$ and $b = \frac{1}{\sigma_n^2} X^T y$ and $X \in \mathbb{R}^{n \times d}, y \in \mathbb{R}^{n \times 1}$ are design matrices. For test point x_*, the predictive distribution of $f_* = f(x_*)$ is the average likelihood of f_* under model $f(x; w)$ with respect to posterior of w.

$$p(f_* \mid x_*, y) = \int p(f_* \mid x_*, w) p(w \mid y) \, dw \tag{2}$$

We can think of the predictive distribution as a linear function $f_* = x_*^T w$ of weights, a normal random variable, and therefore is normal. Therefore, $f_* \mid x_*, X, y \sim \mathcal{N}(x_*^T A^{-1} b, x_*^T A^{-1} x_*)$. The natural extension to Bayesian linear regression is to kernelize it, for example assume a linear model in some feature space $f(x) = \phi(x)^T w$. Instead of considering w as a random variable, we can

An alternative view point of considering w as random variable is to consider the function class that w parameterizes as a random variable.

References

[1] Carl Edward Rasmussen and Williams Christopher. *Gaussian Process for Machine Learning*. MIT Press, 2006.