SL-HW1

development.stdrr

March 2021

1 Question 1

1.1 Top-3 take-home messages

1. Functions defined in an interval $[a, b]$ (like those of linear/polynomial regression) can be interpreted as vectors in the $\mathbb{R}^\infty$ space. Hence, thanks to the knowledge coming from the linear algebra, these functions can be approximated through the notion of orthonormal spaces, i.e. if we can construct a space of non-zero orthogonal vectors, we can find a basis of vectors that approximates the target function as much as desired.

2. Going further, in practice point (1) results in handling dot products between vectors ϕ_j and coefficients β_j; vectors ϕ_j come from a chosen basis, for example from the cosine basis or the Fourier basis that are straightforward to compute; the β_j can be computed through the generalized Fourier expansion, i.e. computing, or approximating, the dot product between the vectors of the basis and the target function; dot products are already easy to compute, but, thanks to the Perceval's theorem, we can further simplify them: the target function $m(\cdot)$ can be approximated directly from the β_j Fourier coefficients. We can control the precision of the approximation choosing the number J of coefficients. Powerful point: no need to recompute the previously computed coefficients as we change J. Even from a finite sample, we can estimate the overall trend of a function.

3. This analysis can be easily extended to functions $m(\cdot, \cdot, ...)$ of higher dimensions.

2 Question 2

The underlying aim of regression splines is again exploiting the properties of linear functions that lead to flexible and powerful yet relatively easy to fit models. Indeed, the domain of the unknown data generating function $m(\cdot)$ is divided into smaller q intervals $(-\infty, \xi_1], [\xi_1, \xi_2], [\xi_2, \xi_3], ..., [\xi_q, +\infty)$ and a polynomial of degree d is fitted in each of these intervals. In more detail, a spline $f(\cdot)$ of degree d can be written as a linear combination such that:

$$f(x) = \sum_{j=1}^{(d+1)+q} \beta_j \cdot g_j(x),$$

having $\mathcal{G}_{d,q} = \left\{g_1(x), ..., g_{d+1}(x), g_{(d+1)+1}(x), ..., g_{(d+1)+q}(x)\right\}$ the set of the g_j transformation functions on the input data x so that $\left\{g_1(x) = 1, g_2(x) = x, ..., g_{d+1}(x) = x^d\right\}$ and $\left\{g_{(d+1)+j}(x) = (x - \xi_j)_+^d\right\}_{j=1}^q$. According to the book *An Introduction to Statistical Learning*, $\mathcal{G}_{d,q}$ can be viewed as a basis for a d-degree polynomial, augmented with one truncated power basis function per knot. Hence, it turns out that regression splines are strictly related to the basis notion, as orthogonal series expansions. Nevertheless, regression splines' basis isn't, in general, orthogonal; this means that, increasing the approximation precision of $m(\cdot)$, i.e. increasing the number q of knots, requires to recompute all the β_j coefficients.