Elements of Statistical Learning
Homework Sheet 3 (Rev 2, 29th April 2021)

Optional: you can hand in solutions to questions 2, 4 and/or 7 for marking if you like. These questions are marked below in blue and with $*$. This is for formative assessment only and not official coursework and does not count towards your final course grade. Hand in solutions via Blackboard on Monday Mar 13th by 1pm. Ensure your name is CLEARLY written on them. If you submit more than one page, ensure that the pages uploaded as a single document and in the right order, please!

1. Show that S_λ (defined by equation (23), slide 27 in Lecture 11) is symmetric and positive semi-definite.

2. (*) *Show how the criterion*

$$\mathrm{RSS}(\theta, \lambda) = (y - N\theta)^T (y - N\theta) + \lambda \theta^T \Omega_n \theta, \tag{1}$$

can give solution

$$\hat{\theta} = (N^T N + \lambda \Omega_n)^{-1} N^T y. \tag{2}$$

by using the ridge regression machinery — hint: using a reparametrisation of ridge. This was from slide 25 of Lecture 11. Assume all of the inverses you need.

3. In kernel density estimation, the kernel $K(x)$ is chosen to satisfy the following properties (i) $K(x) \geq 0$; (ii) $\int K(x)\, dx = 1$ and $\int xK(x)\, dx = 0$. Explain why it would not be desirable to choose a kernel that did not satisfy all of those properties.

4. (*) *Show that the kernel density estimate $\hat{f}_{n,h,K}(x)$ defined on slide 6 of Lecture 12 is a density (i.e. integrates to one).*

5. Let $\hat{f}(x)$ be a kernel density estimator for density $f(x)$ with kernel K. Use methods similar to those on slides 24, 25 and 26 from Lecture 12 to show that

$$\mathrm{bias}\{\hat{f}'(x)\} = h^2 C_3 f'''(x) + \mathcal{O}(h^3). \tag{3}$$

[Assume f is a density that is three times differentiable on $\mathbb{R}$ and the kernel K satisfies $\lim_{x\to\pm\infty} f(x)K(x) = 0$ and that K has at least one derivative.]

6. Define the (basis) functions $\Psi_k(x) = \exp(2\pi i k x)$, for $x \in [0, 1]$ for $k \in \mathbb{Z}$. Show that the set $\{\Psi_k(x)\}_{k\in\mathbb{Z}}$ is orthonormal.

7. (*) *Let $\phi(x)$ be the usual probability density function of the standard normal distribution. Define the function $\psi(x) = \phi''(x)$, the second derivative of the density. Show that $\psi(x)$ satisfies the key wavelet property of $\int_{-\infty}^{\infty} \psi(x)\, dx = 0$. A normalised version of this negative second derivative is known as the Ricker wavelet. Don't confuse the pdf of a standard normal $\phi(x)$ for the typical notation for a father wavelet.*

8. Let $f(x), g(x)$ be two functions with orthogonal series expansions of $f(x) = \sum_\nu f_\nu \xi_\nu(x)$ and $g(x) = \sum_\nu g_\nu \xi_\nu(x)$, where $\{\xi_\nu\}$ is some orthogonal basis for the space of functions we're considering. Show Parseval's relation

$$< f, g >=< F, G >, \tag{4}$$

where $F = \{f_\nu\}_\nu$ and similarly for G and $< f, g >= \int f(x)\overline{g(x)}\, dx$ and $< F, G >= \sum_\nu f_\nu \overline{g_\nu}$. Also prove Plancheral's theorem $||f||^2 = ||F||_\nu^2$, where the norms are defined using the two inner products $< f, g >, < F, G >$, respectively.

9. Given a function, $f(t)$, the continuous wavelet transform (CWT) is

$$\gamma(s, \tau) = \int f(t)\psi^*_{s,\tau}(t)\, dt, \tag{5}$$

where $*$ denotes complex conjugation and the function $f(t)$ can be reconstructed from the CWT by

$$f(t) = \int\int \gamma(s, \tau)\psi_{s,\tau}(t)\, d\tau ds, \tag{6}$$

where the wavelets are generated by a mother wavelet by

$$\psi_{s,\tau}(t) = s^{-1/2}\psi\left(\frac{t-\tau}{s}\right), \tag{7}$$

where s is the scale factor and τ is the translation or location factor. The rth moment of the wavelet is defined by

$$M_r = \int t^r \psi(t)\, dt. \tag{8}$$

A wavelet with p vanishing moments means that $M_r = 0$ for $r = 0, \ldots, p$.

Suppose our wavelet has p vanishing moments, and that $f(t)$ is $(p+1)$-times differentiable. By using a Taylor expansion for $f(t)$ around $t = 0$, show that the wavelet coefficients at scale s are $\gamma(s, 0) = C^* M_{p+1} s^{p+3/2}$, where C^* is some constant. (Knowing the decay/size of wavelet coefficients for functions of a certain smoothness is useful for establishing theoretical results and can be used for characterising function classes.)