Mathematical Foundations of Machine Learning – Spring 2020

Summer assignment

1. [Binary classification] Explain what a ROC curve is. How is it computed for a given trained logistic regression model?
2. [Gaussian Process] Assume that you have time series data of daily sales and you wish to forecast the sales over the next 7 days.
   a. Write explicitly how one uses Gaussian processes to do so.
   b. Assume you have a measure of similarity $0 \le \rho\left(\vec{d}\left(x_i\right), \vec{d}\left(x_j\right)\right) \le 1$, between days (also future days), where $\vec{d}\left(x_i\right)$ is a feature vector for the day x_i. How can one use this measure in the framework of Gaussian process forecasting?
3. [weak l_p spaces] Give an example for a sequence of non-negative numbers that are not in l_2, but are in weak l_2.
4. [Modulus of smoothness] Let $E_N\left(f\right)_p := \inf_{P \in \Pi_N} \left\| f - P \right\|_{L_p[-\pi,\pi]}$, $f \in L_p\left[-\pi,\pi\right]$, $1 \le p \le \infty$, be the degree of approximation by trigonometric polynomials of degree N (not necessarily the Fourier series for $p \ne 2$). Assume that there exists a constant $c > 0$ such that $E_N\left(g\right)_p \le cN^{-1}\left|g\right|_{1,p}$, for any $g \in W_p^1\left[-\pi,\pi\right]$. Prove that $E_N\left(f\right)_p \le \tilde{c}\omega_1\left(f, N^{-1}\right)_p$ for any $f \in L_p\left[-\pi,\pi\right]$.
5. [Besov Spaces] Let $\Omega_1 \subset \Omega_2 \subset \mathbb{R}^n$ be two compact domains. Prove that $B_{p,q}^{\alpha}\left(\Omega_2\right) \subset L_p\left(\Omega_2\right) \cap B_{p,q}^{\alpha}\left(\Omega_1\right)$, for all relevant indices.
6. [Besov spaces over trees] Provide an example for a function $f:\left[0,1\right]^2 \to \mathbb{R}$ and two trees $\mathcal{T}_1, \mathcal{T}_2$ over $\left[0,1\right]^2$, such that for any(!) $\alpha > 0$, $f \in B_{\tau}^{\alpha}\left(\mathcal{T}_1\right)$ and $f \notin B_{\tau}^{\alpha}\left(\mathcal{T}_2\right)$. Here we simply mean $B_{\tau}^{\alpha} := B_{\tau,\tau}^{\alpha}$, with τ any $0 < \tau < \infty$.
7. [Adaptive approximation]. For any $N \ge 1$, find an optimal adaptive N-term piecewise constant approximation to $f\left(x\right) = x^2$ in $L_{\infty}\left[0,1\right]$.
8. [Wavelet Jackson estimate] The proof of the wavelet Jackson inequality for $\sigma_M\left(f\right)_p$ in [6] is for special values of M. The proof ends with a note that that “extending the result for all M is standard”. Prove this extension.
9. [Random Forest] How would you speed up the training of an RF with 5 trees using 10 parallel processors?
10. [DL loss function] Assume that we have a regression DL network $F\left(x,\theta\right)$, where θ are the parameters. Can we train the network using the training set $\left(x_i, y_i\right)_{i \in I}$, with the loss function
    $$L\left(\theta \middle| X\right) = \frac{1}{\# I}\sum_{i \in I}\left|F\left(x_i,\theta\right) - y_i\right| \ ?$$
11. [DL for numerical PDEs] You are given the problem of finding the point source location at time 0, of the 3D wave equation in a the 3D domain $\left[0,1\right]^3$, given only the solution $u\left(x,t\right)$ at fixed time T. Describe how you would create a training dataset for this problem, provide a sketch design for the network architecture and write the loss function. Naturally, one assumes a discretized version of this problem (e.g the input).
12. [Physics aware NN for PDEs] Design a NN architecture and a loss function for the following problem. We need to solve the 2D heat equation

$$\frac{\partial u}{\partial t}(x,t) = \frac{\partial^2 u}{\partial^2 x_1}(x,t) + \frac{\partial^2 u}{\partial^2 x_2}(x,t),$$

in the domain $x = (x_1, x_2) \in [0,1]^2$, $t \in [0,10]$, with initial conditions $u(x,0) = f(x)$, and (compatible) boundary conditions

$$u((0,x_2),t) = g_1(t),\ u((1,x_2),t) = g_2(t),\ u((x_1,0),t) = g_3(t),\ u((x_1,1),t) = g_4(t),\ t \in [0,10]$$

The input to the network is $x = (x_1, x_2) \in [0,1]^2$, $t \in [0,10]$, and the output is a scalar $\tilde{u}(x,t)$. Explain the role of automatic differentiation during training.