Exercises in Convex Optimization
Lecture 11, supplementary exercise

Michel Baes, Patrick Cheridito

December 13, 2018

Universal approximation in Hilbert spaces Let H be a Hilbert space and $G \subseteq H$ be a bounded set in H, say $G \subseteq B[0, R]$. Let x be a point in the closure of $\text{conv}(G)$ and let $\delta > R^2 - ||x||^2$. Prove that for every positive integer n, there exists n point $p_1, \ldots, p_n \in G$ and nonnegative numbers $\lambda_1, \ldots, \lambda_n$ summing up to 1 for which:

$$\left\| x - \sum_{i=1}^{n} \lambda_i p_i \right\|^2 \leq \frac{\delta}{n}.$$

We suggest the following plan for your proof. Let $\epsilon > 0$. Take $x_\epsilon \in \text{conv}(G)$ so that $||x - x_\epsilon|| < \epsilon$ and define $x_1, \ldots, x_M \in G$ so that $x_\epsilon = \sum_{i=1}^{M} \gamma_i x_i$, where the γ_i's are positive and sum up to 1.

1. Suppose that we pick randomly a point from $S_M := \{x_1, \ldots, x_M\}$, so that x_i is taken with probability γ_i. What is the expectation of this random variable?
2. Suppose that we pick randomly n points $p_1, \ldots, p_n$ from S_M, independently (we can therefore pick the same point more than once), with the same distribution as above. What is the expectation of their average $\bar{p} := \frac{p_1 + \cdots + p_n}{n}$?
3. Verify that for every $i \neq j$ we have $\mathbb{E}[\langle x_\epsilon - p_i, x_\epsilon - p_j \rangle] = 0$.
4. Prove that $\mathbb{E}[||x_\epsilon - \bar{p}||^2] \leq (R^2 - ||x_\epsilon||^2)/n$.
5. Since $\delta > R^2 - ||x||^2$, there is some $0 < \gamma < 1$ for which $\gamma\delta \geq R^2 - ||x||^2$. Prove that if $\epsilon \leq \min\left\{\frac{2R}{n}, \frac{\delta(1-\gamma)}{4R}\right\}$, we have $\mathbb{E}[||x - \bar{p}||^2] \leq \frac{\delta}{n}$.
6. Deduce that there exists $y \in H$ that is a convex combination of at most n points of G, and for which $||x - y||^2 \leq \frac{\delta}{n}$.

Note: this result is used in a critical way to relate, in a shallow neural network with sigmoid activation functions, the number of neurons and the optimal approximation accuracy of this network.

The formal statement is as follows.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function (the function to be approximated by a neural network) that has a Fourier transform $\hat{f}$ defined on $\mathbb{R}^n$. Assume that the functions $\omega \mapsto \exp(i\omega^T x)\hat{f}'(\omega)$ are integrable for every $x \in \mathbb{R}^n$ and that these integrals are uniformly bounded by a constant C. Let $\phi : \mathbb{R} \to [0, 1]$ be a measurable function for which $\lim_{t\to\infty} \phi(t) = 1$ and $\lim_{t\to-\infty} \phi(t) = 0$ (this ϕ is the activation function of the neurons). Let $f_n(x; A, b, c) := \sum_{i=1}^{n} c_i\phi(\langle a_i, x\rangle + b_i) + c_0$ for a real matrix $A = [a_1, a_2, \ldots, a_n]$ and vectors b, and c of appropriate size (this $f_n(\cdot; A, b, c)$ is the function implemented by a shallow neural network with n nodes and parameters A, b, c). Finally, let μ_r be an arbitrary probability measure on the ball $B(0, r)$ for some fixed radius $r > 0$.

Then for every $n \geq 1$ there exist A, b, c such that

$$\int_{B(0,r)} ||f_n(x; A, b, c) - f(x)||^2 d\mu_r(x) \leq \frac{4r^2C^2}{n}.$$

Here is quite a rough proof sketch.

We can assume that $f(0) = 0$ as it suffices to adjust c_0 appropriately.

Let G_F be the set of functions of the form $cF(\langle a, x\rangle + b)$ for $|c| \leq 2C$. By the result proved in the above exercise, it suffices to verify that f is in the closure of $\mathrm{conv}(G_\phi)$ (and to prove that we can take δ as small as $4r^2C^2$, but this is actually pretty elementary).

That statement is first proved for $G_{\cos}$ by using the inverse Fourier transform of $\hat{f}$. Denoting by "step" the function $\mathbf{1}_{\{t\geq 0\}}$, we can show that $G_{\cos}$ is in the closure of $\mathrm{conv}(G_{\mathrm{step}})$ by representing the function "cos" as a limit (in the $L^2(\mu_r)$ sense) of functions in $\mathrm{conv}(G_{\mathrm{step}})$. Finally, we can show that the function "step" can be written as the limit (also in the $L^2(\mu_r)$ sense) of a sequence of functions in $\mathrm{conv}(G_\phi)$.

`Lecture_12/HilbertConvexApproximation`