Lecture notes from
Models and Numerical Methods

https://github.com/Grufoony/Physics_Unibo

Contents

1 Resume of measure theory

We need to define a mathematical model that generates sequences from an *alphabet* $\mathcal{A}$, which can be any finite set. We will denote both set of finite and infinite sequences as $\mathcal{A}^* = \cup_{n\in\mathbb{N}}\mathcal{A}^n$ and $\mathcal{A}^{\mathbb{N}}$. Now we can define a sequence, or a *word*, $\omega \in \mathcal{A}^n$ and denote with $|\omega| = n$ its length. In particular, we will use the notation $\omega_i^j = (\omega_i, \dots, \omega_j)$. We can also take $\mathcal{A}^{\mathbb{Z}}$ as two-sided alphabet.

Definition 1. *A **measurable space** $(\Omega, \mathcal{F})$ is (usually) defined by a compact metric space Ω and a σ-algebra $\mathcal{F}$.*

We will denote the canonical cylinder on Ω as $[a_1^n] = \{y \in \Omega \mid y_1 = x_1, \dots, y_n = x_n\}$. To figure out that this is actually a cylinder, let's pretend to take (r, ϕ, h) cylindrical coordinates, fixing the radius $r = r_0$, letting the angle and the heigh free.

Our space has a topology, so we can take $\mu \approx$ m (metric) absolutely continuous w.r.t. the Lebesgue measure on $\Omega = \mathbb{R}^n$. So it exists $\varphi \in \mathbb{L}^1(m)$ such that $\mu(f) = \int dm f(m)\varphi(m)$. Consider now the function

$$g_{\mathcal{A}}(z, z') = \begin{cases} 1 & z = z' \\ 0 & z \neq z' \end{cases} \qquad \forall z, z' \in \mathcal{A}$$

Taking $x, y \in \Omega$ infinite sequences it is possible to prove that

$$\widetilde{d}(x, y) = \sum_{n=1}^{\infty} 2^{-n} g_{\mathcal{A}}(x_n, y_n)$$

is a metric over Ω. Taking $x^{(n)} \in \Omega$ sequence of infinite sequences, given $0 < \lambda = \frac{1}{|A|} < 1$, we have that $d(x, y) = \lambda^{n(x,y)} \quad \forall x, y \in \Omega$ is also a metric over Ω, with $n(x, y) = \min\{k | x_k \neq y_k\}$. Moreover, d and $\widetilde{d}$ define the same topology. The open balls are, $\forall x \in \Omega, \quad r > 0$

$$\mathcal{B}(x, r) = \{y \in \Omega \mid d(x, y) \leq r\} = \left\{ y \in \Omega \mid x_k = y_k \ \forall \ 1 \leq k \leq \frac{\ln r}{\ln \lambda} \right\}$$

Definition 2. *$\mathcal{F}$ is a **Borel σ-algebra** if is a set of subsets of Ω such that $\Omega \in \mathcal{F}$*

So a σ-algebra is actually a collection of all measurable sets.

2 Stochastic Processes

Definition 3. *A **stochastic process** is an infinite sequence of random variables X_n with values in $\mathcal{A}$ defined by the k^{th} order joint distribution:*

$$\mu_k\left(a_1^k\right) = \mathbb{P}\left(X_1^k = a_1^k\right) \quad a_1^k \in \mathcal{A}$$

We need also a consistency condition:

$$\mu_t\left(a_1^t\right) = \sum_{a_0 \in \mathcal{A}} \mu_{t+1}\left(a_0^t\right) = \sum_{a_{t+1} \in \mathcal{A}} \mu_{t+1}\left(a_1^{t+1}\right)$$

Equivalently, we can define a stochastic process through the conditional probability

$$\mu\left(a_t | a_1^{t-1}\right) = \frac{\mu_t\left(a_1^t\right)}{\mu_{t-1}\left(a_1^{t-1}\right)}$$

The μ_k are called **marginals** and, in order to be a probability, they must satisfy the normalization condition

$$\sum_{a_1^k \in \mathcal{A}} \mu_k\left(a_1^k\right) = 1$$

We notice that this sum is exponentially growing in k, so it's impossible to approximate the measure.

Definition 4. *A stochastic process is **stationary** if*

$$\mu\left(a_1^k\right) = \mu\left(a_{t+1}^{t+k}\right) \quad \forall a_1^\infty \in \mathcal{A}^{\mathbb{N}}$$

Definition 5. *An **information source** is a stationary, ergodic, stochastic process.*

Definition 6. *A process or a source is a **shift-invariant Borel probability measure** μ on the topological space $\mathcal{A}^{\mathbb{Z}}$ of doubly-infinite sequences $x = \{x_n\}_{n\in\mathbb{Z}}$, drawn from a finite (i.e. countable) alphabet $\mathcal{A}$*

Furthermore, it is trivial that we can write any standard cylinder as

$$\left[x_1^t\right] = \sqcup_{a\in\mathcal{A}}\left[x_1, \ldots, x_t, a\right]$$

It's easy to check that

$$\mu \in \mathcal{P}_I\Omega \mid \mu \circ \sigma^{-1} = \mu \Leftrightarrow \sum_{a\in\mathcal{A}} \mu_{t+1}\left(a, x_1, \ldots, x_t\right) = \mu_t\left(x_1^t\right)$$

Neural networks are heuristically approximating μ.

Theorem 1 (Kolmogorov representation theorem)**.** *If $\{\mu_n\}$ is a sequence of measure defining a process then there is a unique Borel probability measure μ on $\mathcal{A}^\infty$ such that, $\forall k \geq 1$ and $\forall \left[a_1^k\right]$ cylinder*

$$\mu\left(\left[a_1^k\right]\right) = \mu_k\left(a_1^k\right)$$

2.1 Markov's Models

2.2 Hidden Markov's Models