RECURRENT NEURAL NETWORKS

JEAN-MARTIN ALBERT

1. The basic problem

Let $U = \mathbb{R}^n$ and $V = \mathbb{R}^m$ be vector spaces, and consider a function $f : U \to W$. In theory, f here can be any set theoretic function, but for our purpose, we will assume that f is at the very least square integrable. In general, we consider a class $\mathcal{C}$ of functions $U \to V$, and try to find an element $g \in \mathcal{C}$ which is as close as possible to f. In order to do so, we need a notion of distance between functions. There are many possible choices, but some of the most popular are the L^p norms and the L^∞ norm.

While the class $\mathcal{C}$ van be virtually any class of functions, we usually choose a class that is somewhat definable, in the sense that there is a (differentiable) function $F : U \times \mathbb{R}^k \to V$ such that $\mathcal{C} = \{F(x, w) : w \in \mathbb{R}^k\}$. Such classes of functions have the advantage of being equicontinuous.

Let $f(x) = x(x-1)(x+1)$. As an example, we will try to fit a linear function $\ell(x) = ux + v$ to f. The class $\mathcal{C}$ is defined by $\mathcal{C} = \{ux + v : u, v \in R\}$. For an element $g(x, u, v) \in \mathcal{C}$, we have

$$f(x) - g(x, u, v) = x^3 - x - ux + v$$

From this we get

$$\begin{aligned}(f(x) - g(x, u, v))^2 &= (x^3 - (1-u)x + v)^2 \\ &= x^6 - 2(1-u)x^4 + 2vx^3 + (1-u)^2x^2 - \\ &\quad 2(1-u)vx + v^2\end{aligned}$$

If we compute $\int_{-1}^1 (f(x) - g(x, u, v))^2$, we get

$$\begin{aligned}\int_{-1}^1 (f(x) - g(x; u, v))^2 &= \int_{-1}^1 (x^3 - (1-u)x + v)^2 \\ &= \int_{-1}^1 x^6 - 2(1-u)x^4 + 2vx^3 + (1-u)^2x^2 - 2(1-u)vx + v^2 \\ &= \frac{x^7}{7} - 2v\frac{x^4}{4} + (1-u)^2\frac{x^3}{3} - 2(1-u)v\frac{x^2}{2} + v^2x \\ &= \frac{1}{7} - 2v\frac{1}{4} + (1-u)^2\frac{1}{3} - 2(1-u)v\frac{1}{2} + v^2 \\ &\quad + \frac{1}{7} - 2v\frac{1}{4} + (1-u)^2\frac{-1}{3} - 2(1-u)v\frac{1}{2} - v^2 \\ &= \frac{2}{7} - (2-u)v + v^2 = \ell(u, v)\end{aligned}$$

Date: August 17, 2017.

and note that the last line above is a function of u and v alone, and we have to minimize it.

2. Dealing with sequences