Appunti di Analisi Funzionale

Github Repository: `0xke/appunti/AnalFun`

Primo semestre, 2025 - 2026, prof. Antonio Edoardo Segatti

0.1 Neural Networks

0.1.1 MLP

Definizione 0.1.1: Multi Level Perceptron

Una Multi-Level Perceptron (**MLP**) è una mappa $\varphi : \mathbb{R}^d \to \mathbb{R}^{n_L}$ tale che $\varphi(x_0) = x_L$ e

$$x_l = \rho_L(A_l x_{l-1} + b_l) \quad \forall l = 1, \ldots, L \tag{0.1.1}$$

con $\rho_l : \mathbb{R} \to \mathbb{R}$ componente per componente, funzione di attivazione.

La precedente è una rete "feedforward".

Definizione 0.1.2: ResNet

Prendendo $\rho_l(x) = x_{l-1} + \rho(x)$ in (0.1.1) viene una rete neurale "resuduale" di cui un'implementazione è la rete *ResNet* e per l'ottimizzazione può funzionare meglio, nonostante per l'approssimazione non cambia molto.
Inoltre può essere vista come discretizzazione di

$$\dot{x}(t) = \rho(A(t)x(t) + b(t))$$

Definizione 0.1.3: Recurrent NNs

Dati input $y_1, y_2, \dots \in \mathbb{R}^{n_{in}}$, modifichiamo l'operazione (0.1.1) in

$$x_l = \rho(A_x x_{l-1} + A_y y_l) \quad l = 1, 2, \ldots$$

con $x_0 \in \mathbb{R}^N$. Possono essere usate ad esempio per risolvere

$$\begin{cases} \overline{x}(t) = F(x(t), y(t)) \\ x(0) = x_0 \end{cases}$$

Capitolo 1

Errors

Ci sono 3 errori che si possono verificare in ambito di questo tipo di matematica applicata:

1. Approssimazione
2. Generalizzazione
3. Training

Principlamente questo corso si occuperà principalmente di comprendere l'errore dovuto all'approssimazione, parlando meno degli altri due errori.

Siano X, Y due insiemi, con μ una misura su X, $\|\cdot\|_Y$ una norma su Y e $G : X \to Y$ una funzione. Possiamo allora definire una **loss function**

$$\mathcal{L}(\Phi) = \int_X \|G(x) - \Phi(x)\|_Y^2 \, d\mu(x)$$

dove $\Phi \in \mathcal{C}$ una classe di funzioni considerate per l'approssimazione

Non potendo avere misuramenti di G per infiniti valori, si prende in realtà un sample $\{x_i\}_{i=1}^N$ di input e $\{y_i = G(x_i)\}_{i=1}^N$ per cui la loss calcolabile è la **empirical loss**

$$\tilde{\mathcal{L}}(\Phi) = \sum_{i=1}^n w_i \|y_i - \Phi(x_i)\|_Y^2$$

con i w_i pesi. Chiamiamo Φ_{min} e $\tilde{\Phi}_{min}$ le funzioni che sarebbero i minimi su $\mathcal{C}$ delle due loss rispettivamente.

In pratica, un'algoritmo di ottimizzazione è usato per minimizzare $\tilde{\mathcal{L}}$ e Φ_{comp} è l'approssimazione calcolata. Allora

$$\|G - \Phi_{comp}\| \leq \underbrace{\|G - \Phi_{min}\|}_{\text{approx error}} + \underbrace{\|\Phi_{min} - \tilde{\Phi}_{min}\|}_{\text{generalization error}} + \underbrace{\|\tilde{\Phi}_{min} - \Phi_{comp}\|}_{\text{training error}}$$

In particolare cosa vogliamo arrivare a dimostrare noi è il *universal approximation theorem*, ossia un teorema che dia le ipotesi per poter avere che l'errore tende a zero.

1.1 Setting

Sia $K \subseteq \mathbb{R}^d$ un compatto. Sia $C(K)$ l'insieme delle funzioni continue $K \to \mathbb{R}$. Sia ρ una funzione di attivazione continua. Sia $\mathcal{M}$ la famiglia

$$\mathcal{M} = \{\mu \text{ misura relativa di Borel su } K \text{ con variazione totale finita } \}$$

ossia il duale topologico di $C(K)$

Teorema 1.1.1: Template of a Universal Approximation Theorem

Under some conditions on ρ,

$$MLP(\rho, d, \sim) \text{ is dense in } C(K)$$

By Riesz Theorem $\forall L \in C(K)'$, $\exists \mu \in \mathcal{M}$ such that $Lf = \int_K f\, d\mu$

Definizione 1.1.1: Discriminant

We say that $f \in C(K)$ is **discriminant** if

$$\int_K f(a^T x + b)\mu(dx) = 0 \quad \forall a \in \mathbb{R}^d, \forall b \in \mathbb{R}$$

We will prove that on condition that makes the theorem true is to have ρ be *discriminant*. An easier constraint on ρ, is have ρ not be polynomial.

Esercizio 1.1.1

1. Argue that the thesis of universal approx theorem can't be true if ρ is a polynomial
2. Conclude that a polynomial can't be discriminant
3. Show 2. with the definition