

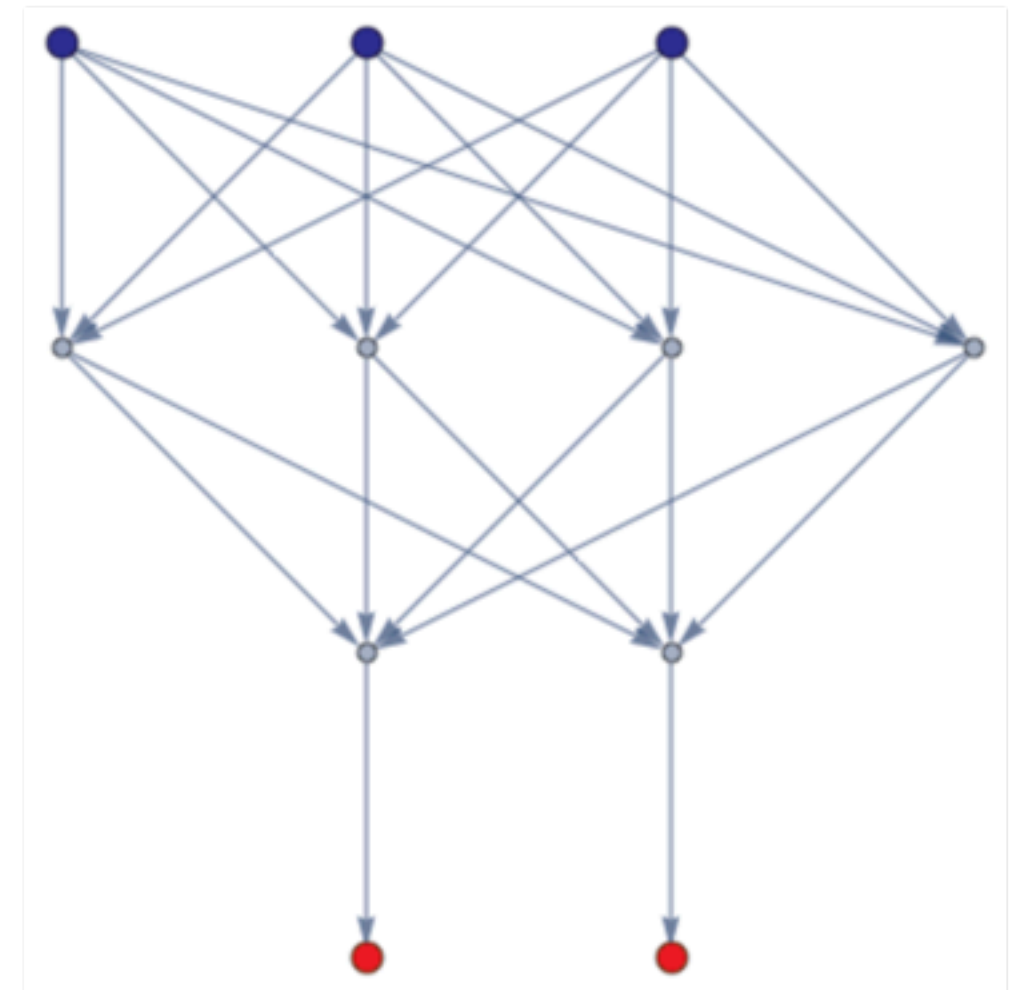


Wprowadzenie do wykorzystania pakietów nnet i caret

Michał Ramsza

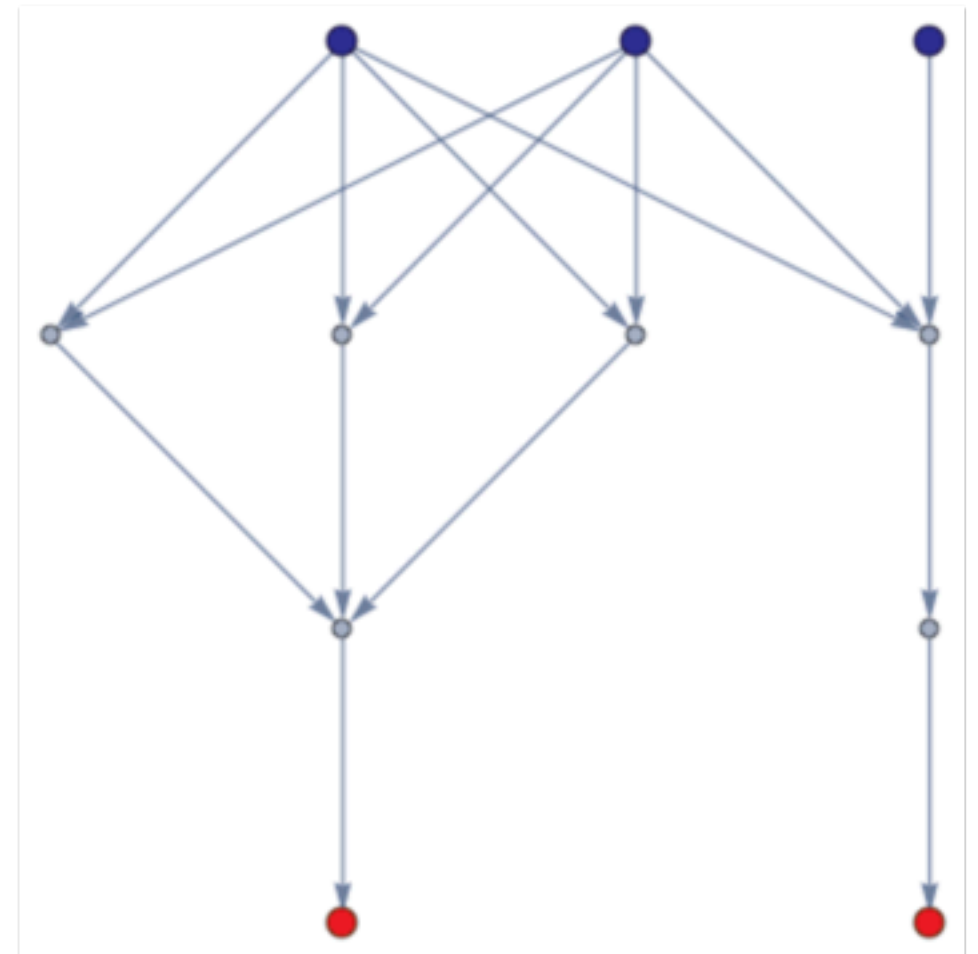
Sieć neuronowa / struktura

- Niech dany będzie graf skierowany $G = (V, E)$, gdzie V jest zbiorem wierzchołków a E jest zbiorem krawędzi.
- Zakładamy, że w grafie nie ma cykli.
- W zbiorze wierzchołków V wyróżniamy podzbiór I (wejścia) oraz podzbiór O (wyjścia). Te wierzchołki nie realizują żadnych funkcji.
- Pozostałe wierzchołki tworzą podzbiór N .
- Dla dowolnego $n \in N$ przez $\text{pre}(n)$ oznaczamy zbiór wierzchołków $v \in V$ spełniających $(v, n) \in E$.



Sieć neuronowa / struktura

- Niech dany będzie graf skierowany $G = (V, E)$, gdzie V jest zbiorem wierzchołków a E jest zbiorem krawędzi.
- Zakładamy, że w grafie nie ma cykli.
- W zbiorze wierzchołków V wyróżniamy podzbiór I (wejścia) oraz podzbiór O (wyjścia). Te wierzchołki nie realizują żadnych funkcji.
- Pozostałe wierzchołki tworzą podzbiór N .
- Dla dowolnego $n \in N$ przez $\text{pre}(n)$ oznaczamy zbiór wierzchołków $v \in V$ spełniających $(v, n) \in E$.



Sieć neuronowa / obliczenia

- Wszystkie wierzchołki $n \in N$ realizują funkcje postaci

$$f_n(x_n) = \theta_n(u_n(x_n)),$$

gdzie θ to funkcja aktywacji a u_n to funkcja agregacji.

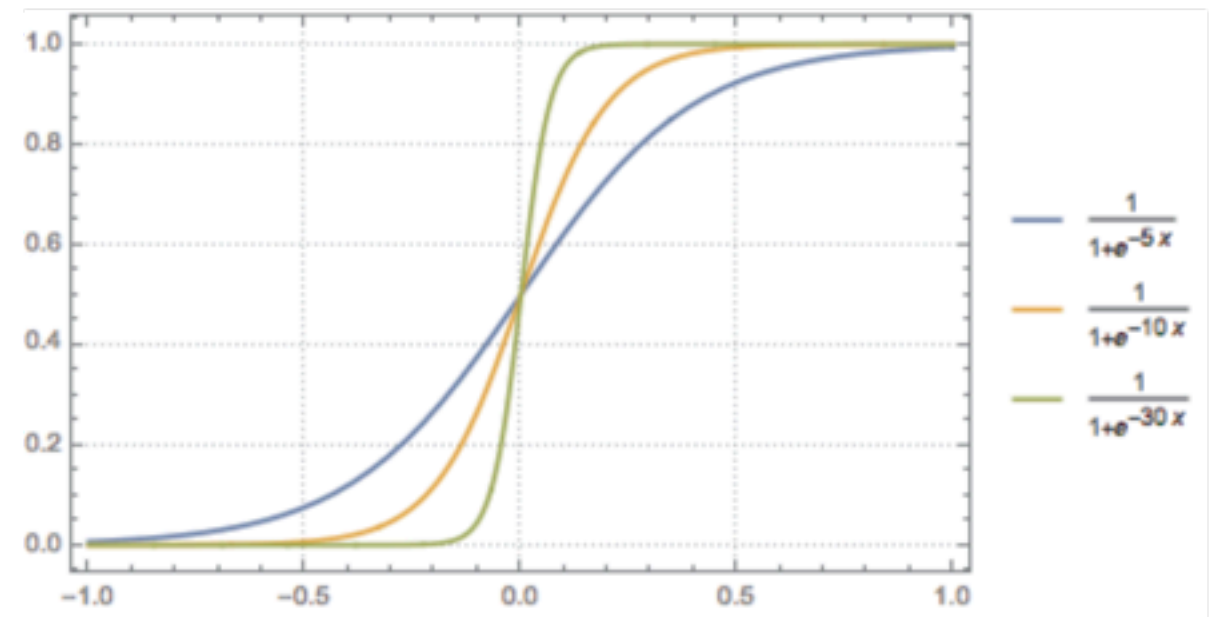
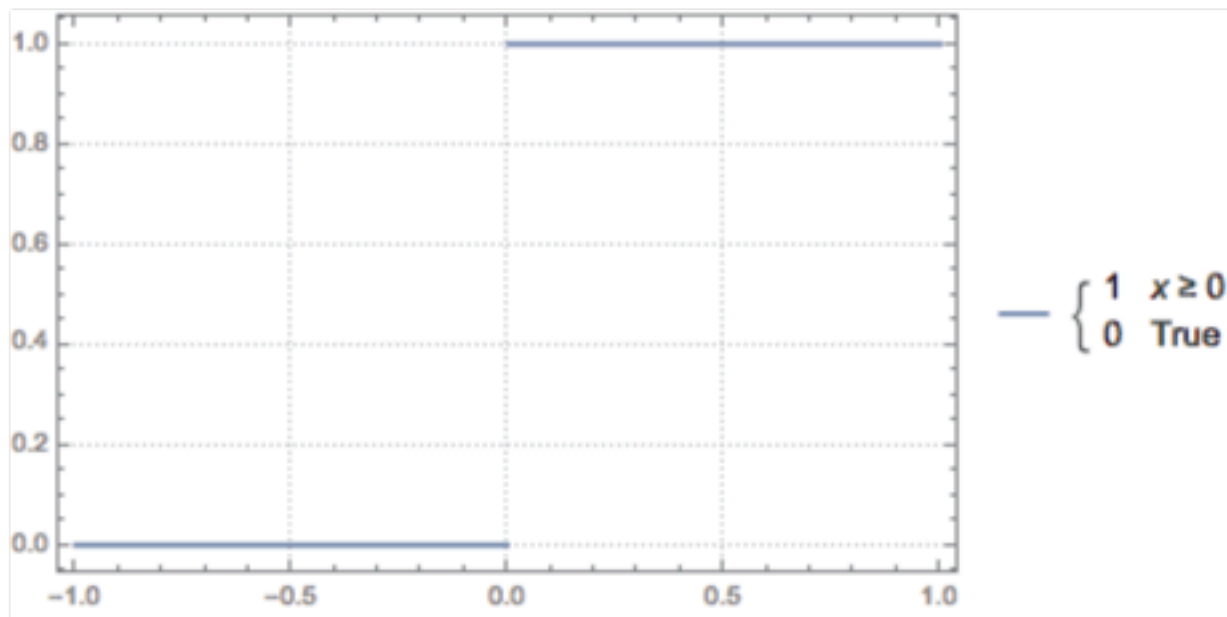
Wektor $x_n = (\dots, x_{nk}, \dots)$ dla dowolnego $k \in \text{pre}(n)$

spełnia $x_{nk} = f_k(x_k)$.

- Dla dowolnego $n \in N$ funkcje f_n są parametryzowane wektorem parametrów w_n . Wszystkie parametry w sieci tworzą wektor w .

Sieć neuronowa / obliczenia / typowe funkcje

- Typowe funkcje agregujące są proporcjonalne do iloczynów skalarnych i metryk.
- Typowe funkcje aktywacji to np. funkcja Heaviside'a czy funkcja logistyczna.



Sieć neuronowa / uczenie

□ Zakładamy, że dana jest próba $\{(x_j, y_j)\}_{j=1}^J$, gdzie x_j to dane wejściowe a y_j to odpowiedź sieci.

□ Dla dowolnych danych wejściowych x_j sieć oblicza wartości wyjściowe \bar{y}_j . Na wektorach y i \bar{y} zdefiniowana jest funkcja kosztów C , której wartość przy ustalonej próbie zależy jedynie od wektor parametrów w , $C(w)$.

□ Zadanie uczenia sieci neuronowej to zadanie minimalizacji kosztu

$$\min_w C(w).$$

□ Typowe funkcje kosztu to koszt średniokwadratowy czy entropia krzyżowa.

□ W celu zwalczania przeuczenia, do funkcji kosztu dodaje się czasami składnik regularyzujący. W najprostszej wersji może to być następujący składnik

$$C(w) + \lambda \langle w | w \rangle,$$

gdzie parametr λ odpowiada za siłę regularyzacji.