

Свёрточные ИНС, глубокие ИНС. Методы ИИ в задачах анализа текстов

Зубюк Андрей Владимирович
zubjuk@physics.msu.ru

<http://NeuroFuzzy.Phys.MSU.ru>

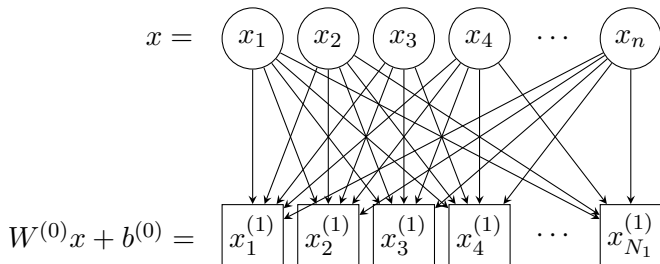
ИНС с 1 скрытым нелинейным слоем — перцептрон Розенблатта.

F. Rosenblatt, “The perceptron. a perceiving and recognizing automaton (project para)”,
Cornell Aeronautical Laboratory, Inc., Tech. Rep. 85-460-1, Jan. 1957

n -мерный случай

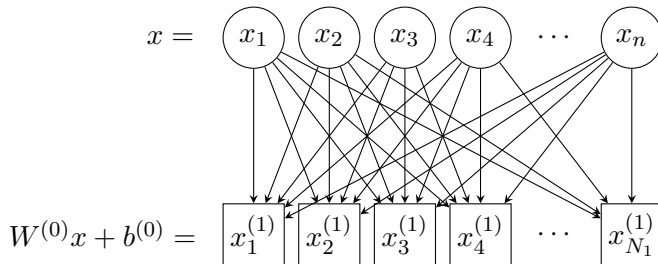
- ▶ Количество нейронов на скрытом слое — k^n .
- ▶ Для изображения n — это количество пикселей ($\sim 10^6$ — мегапиксель). Следовательно, необходимо $\sim 2^{10^6}$ **нейронов**, что больше, чем количество атомов во вселенной.

Перцептрон Розенблатта Vs глубокая ИНС

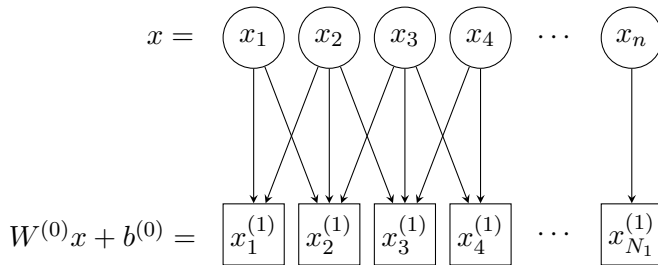


- ▶ При одном скрытом слое точность аппроксимации достигается увеличением «ширины» слоя (количеством нейронов в нём). При этом количество параметров ИНС (весов) растёт линейно с ростом «ширины» слоя.
- ▶ При 2 и более скрытых слоях рост «ширины» слоёв приводит к квадратичному росту количества параметров (весов). Таким образом, одновременное увеличение количества слоёв («глубины» ИНС) и их «ширины» крайне затратно — в процессе обучения понадобится подобрать значения очень большого числа параметров.
- ▶ Имеем два принципиально разных подхода к увеличению сложности ИНС:
«уширение» единственного слоя и увеличение «глубины».

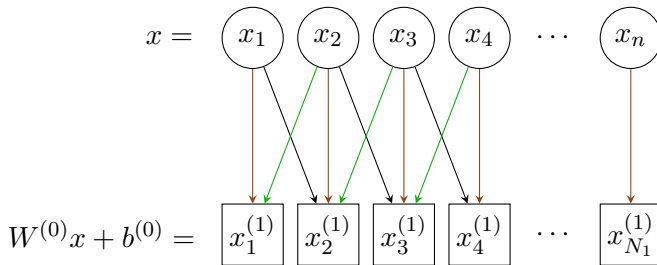
Полносвязные Vs «разреженные» слои ИНС



Полносвязные Vs «разреженные» слои ИНС



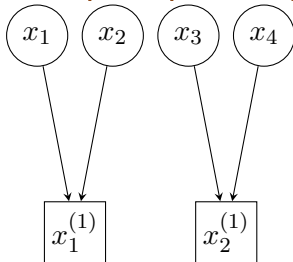
Вариант «разреженного» слоя — свёрточный слой



- ▶ Смещение $b^{(0)} = 0$.
- ▶ Веса связей, ведущих к $x_1^{(1)}$, $x_2^{(1)}$ и т.д. одинаковы: w_1 , w_2 , w_3 .
- ▶ Совокупность весов w_1 , w_2 , w_3 , ... называют **ядром свёртки**.

Другие операции, применяемые в свёрточных ИНС

Понижение размерности путём агрегирования — пулинг (pooling)



В качестве агрегирующей операции часто используется

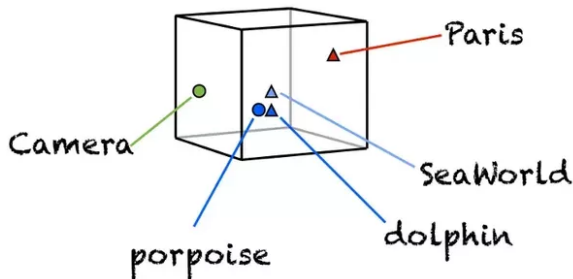
- ▶ **среднее арифметическое** (в ранних работах по свёрточным ИНС) или
- ▶ **max** (в современных работах).

Функция активации ReLU

$$\text{ReLU}(x) = \max\{0, x\} = \begin{cases} 0, & x < 0, \\ x, & x \geq 0. \end{cases}$$

В некотором смысле является простейшей нелинейной функцией: состоит из 2-ух линейных фрагментов, дифференцируема, производная равна 0 (в отрицательной области) или 1 (в положительной).

- ▶ **Мешок слов** (bag of words) — подход, при котором анализируются только частоты появления слов в тексте, порядок слов не учитывается. В этом случае признаки текста — это частоты слов.
- ▶ Для более тонкого анализа текстов необходимо рассматривать как признаки сами слова. Слово — признак категориального типа. Их можно перенумеровать, но арифметические операции над их номерами не являются осмысленными.
- ▶ Для применения метода разделения гиперплоскостью, ИНС и многих других методов в задачах анализа текстов необходимо преобразовать слова в векторы признаков, т.е. построить **отображение** множества слов в признаковое пространство — **эмбединг** (embedding).



- ▶ Каждому слову ставится в соответствие вектор: Word $\xrightarrow{\text{to} - 2}$ Vec.
- ▶ Векторы, соответствующие словам, подбираются так, чтобы словам, имеющим близкий смысл, соответствовали близкие векторы.
- ▶ Показателем близости смысла считается употребление в схожих контекстах. Контекст слова понимается тривиальным образом — как несколько слов, окружающих выбранное (например, 5 предшествующих слов и 5 последующих).
- ▶ Подбор векторов осуществляется методами машинного обучения на основе корпуса текстов.