

Специальные архитектуры нейросетей

Виктор Китов
victorkitov.github.io

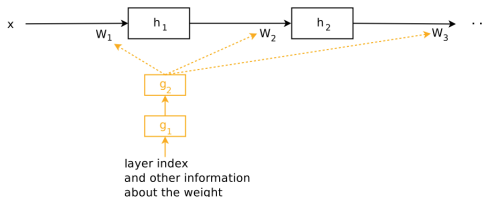
Особые применения

Нейросети (как и др. модели) могут предсказывать не точки, а плотности распределений:

- предсказывать гистограмму из K столбцов.
- предсказывать параметры семейства, например μ, Σ для $y|x \sim \mathcal{N}(\mu(x), \Sigma(x))$

Гиперсеть²

- Гиперсеть (hypernetwork) генерирует параметры θ для другой сети $F_\theta(x)$:



- Настройка: $\theta = g(\text{layer}, \dots)$, $\nabla \mathcal{L}(F_\theta(x), y)$ меняет параметры только гиперсети $g(\cdot)$.
- Пример: для разных подклассов задачи использовать разные параметры основной сети. Гиперсеть их подстраивает, в зависимости от подзадачи¹.

¹<https://arxiv.org/pdf/1906.00695.pdf>

²<https://arxiv.org/pdf/1609.09106.pdf>

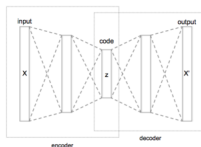
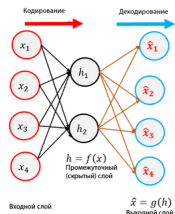
Автокодировщик

- Автокодировщик - архитектура, пытающаяся восстанавливать вход:

$$\hat{x} = g(f(x)), \quad \hat{x} \approx x$$

- $f(x)$: кодировщик
- $h = f(x)$: компактное промежуточное представление
- $g(h) = g(f(x))$: декодировщик

Недоопределенный автокодировщик



Недоопределенный автокодировщик

Недоопределенный автокодировщик (Undercomplete Autoencoder)

$$\dim(h) < \dim(x)$$

$$\mathcal{L}(\hat{x}, x) = \|\hat{x} - x\|_2^2$$

f и g могут содержать несколько слоев.

Применения автокодировщика

Применения:

Применения автокодировщика

Применения:

- компрессия данных: $\dim(h) < \dim(x)$
- снижение размерности, анализ данных в 2D, 3D.
- извлечение информативных признаков $x \rightarrow h$ для др. задач
 - регрессия, классификация, ранжирование и др.
- инициализация первых слоев др. архитектуры кодировщиком $f(x)$
- фильтрация шума: сложно воспроизвести нетипичность
- детекция аномалий: там, где $\|\hat{x} - x\|$ велико.

Другие автокодировщики

- **Разреженный автокодировщик (sparse autoencoder)**
 - может иметь $\dim(h) \geq \dim(x)$
 - компактность h обеспечивается регуляризацией

$$\mathcal{L}(\hat{x}, x) = \|\hat{x} - x\|_2^2 + \lambda R(h)$$

например, $R(h) = \|h\|_1$ или $R(h) = \|h\|_2^2$

Другие автокодировщики

- **Разреженный автокодировщик** (sparse autoencoder)
 - может иметь $\dim(h) \geq \dim(x)$
 - компактность h обеспечивается регуляризацией

$$\mathcal{L}(\hat{x}, x) = \|\hat{x} - x\|_2^2 + \lambda R(h)$$

$$\text{например, } R(h) = \|h\|_1 \text{ или } R(h) = \|h\|_2^2$$

- **Фильтрующий автокодировщик** (denoising autoencoder)
восстанавливает x по $x + \varepsilon$:

$$\hat{x} = g(f(x + \varepsilon)), \quad \hat{x} \approx x$$

- примеры шума: Гауссов, зашумление 0 или 0/1.

Другие автокодировщики

- Сжимающий автокодировщик³ (contractive autoencoder) выучивает h , устойчивое к малым изменениям

$$\mathcal{L}(\hat{x}, x) = \|\hat{x} - x\|_2^2 + \lambda \sum_i \left\| \frac{\partial h_i}{\partial x} \right\|_2^2$$

- $h(x)$ изменяется при изменении $x \in X$ из-за $\|\hat{x} - x\|_2^2$
- но не изменяется при прочих изменениях ортогонально X
 - как в PCA - изменения представления только вдоль аппрокс. подпространства

³Explicit Invariance During Feature Extraction.

Заключение

- Нейросеть может предсказывать распределения, а не точечные ответы.
- Гиперсеть предсказывает параметры для другой сети.
- Автокодировщик сжимает данные и получает информативные признаки.