

Метрический анализ пространства параметров глубоких нейросетей.

Насыров Э.Р.

МФТИ

15 марта 2023 г.

Визуализация нейронов сети методом бутстрэп.

Модель двуслойной нейросети:

$$f(x) = \sigma(W_1^T \cdot \sigma(W_2^T x + b_1) + b_2)$$

Модель RNN:

$$h_t = \sigma(W \cdot h_{t-1} + V \cdot x_t),$$

$$s_{t+1} = \tanh(w_o^T \cdot h_t)$$

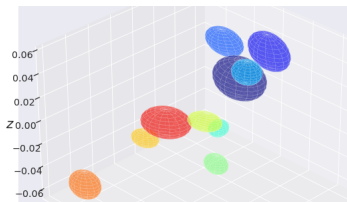


Рис.: Доверительные области нейронов сети 10-3-20.

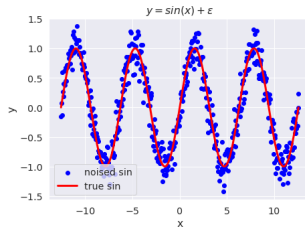


Рис.: Зашумленный sin.

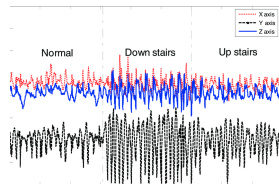


Рис.: Показания акселерометра во время ходьбы.

Содержание

- 1 Введение
- 2 Методы улучшения обучения
- 3 Мотивация
- 4 Постановка задачи
- 5 Используемые модели
- 6 Данные
- 7 Планируемые результаты

Введение

- Высокоразмерные данные - видео, звук - избыточны.
- Модели тяжело обучаются на избыточных данных, часто переобучаются.
- Нужно бороться с избыточностью и переобучением.

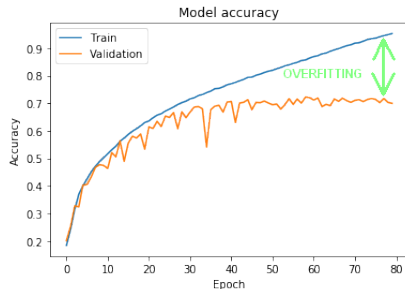


Рис.: Переобучение.

Методы улучшения обучения

- Методы снижения размерности входных данных:
 - ▶ PCA
 - ▶ Quadratic Programming Feature selection
 - ▶ Neural Autoencoders
- Выбор оптимальной структуры модели:
 - ▶ Optimal Brain Surgeon
 - ▶ Correlational analysis
 - ▶ Weights freezing

Мотивация

- В предыдущих работах параметры модели w рассматриваются как отдельные скаляры.
- Не учтена простая структура нейросети - композиция линейных и простых нелинейных функций.
- Составной блок нейросети:

$$y = \sigma(Wx + b),$$

$$y, b \in \mathbb{R}^m, x \in \mathbb{R}^n, W \in \mathbb{R}^{m \times n}, \sigma : \mathbb{R} \rightarrow \mathbb{R}.$$

- В нашей работе будут исследованы не отдельные элементы W , а ее строки w_i , которые в нейросети называют *нейронами*:

$$W = \begin{pmatrix} w_1^T \\ \vdots \\ w_m^T \end{pmatrix}.$$

Постановка задачи

- Дан временной ряд $x = [x_1, \dots, x_N]^T$, $x_i \in \mathbb{R}$, ширина окна n .
- Точка $x_t = [x_t, \dots, x_{t+n-1}]^T$ - точка фазовой траектории временного ряда в траекторном пространстве $\mathbb{H}_x \subset \mathbb{R}^n$.
- Предположение: точка фазовой траектории распределена *нормально* в фазовом пространстве.
- Тогда параметры обученной нейросети будут случайными.

Постановка задачи

- Для каждого вектора-параметра w_i оцениваем его матожидание $e_1 = Ew_i$ и ковариационную матрицу $D_i = \text{cov}(w_i)$.
- Для каждого вектора вычисляем его 90% вероятностную область и траекторном пространстве.
- Проецируем на 2-х мерное (3-х) мерное пространство.
- Визуализируем результаты.

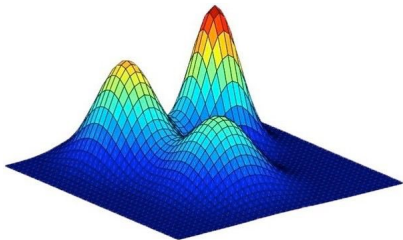


Рис.: Смесь гауссианов трех 2-х мерных векторов.

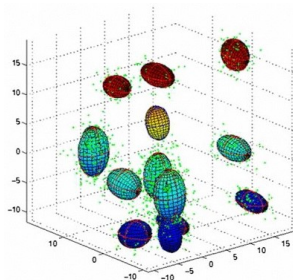


Рис.: Доверительные области 3-х мерных векторов.

Формальная постановка задачи

- Решается задача авторегрессионного декодирования.
- Обозначим множество всех одномерных временных рядов через \mathbb{S} :

$$\mathbb{S} = \bigcup_{n=1}^{+\infty} \{[s_1, \dots, s_n] \in \mathbb{R}^n\}.$$

- Прогностическая модель $f^{\text{AR}} : \mathbb{S} \rightarrow \mathbb{R}$ предсказывает следующее значение временного ряда по предыдущим.
- Модель $f = f(w, s)$, $w \in \mathbb{W}$, $s = [s_1, \dots, s_t] \in \mathbb{R}^t$ выбирается из некоего параметрического семейства. Параметры модели выбираются таким образом, чтобы минимизировать функцию ошибки $S = S(w|s, f)$:

$$w^* = \arg \min_{w \in \mathbb{W}} S(w|s, f).$$

В работе будет использоваться функция ошибки MSE, то есть

$$S(w|s, f) = \sum_{t=h+1}^T (s_t - \hat{s}_t)^2.$$

Используемые модели

- Нелинейный PCA:

$$f(x) = \sigma(w^T \cdot \sigma(W^T x + b_1) + b_2)$$

$$x \in \mathbb{R}^h, W \in \mathbb{R}^{h \times d}, w \in \mathbb{R}^d : w^T w = 1, WW^T = I.$$

- RNN:

$$h_t = \sigma(W \cdot h_{t-1} + V \cdot x_t),$$

$$s_{t+1} = \tanh(w_o^T \cdot h_t)$$

Данные

- Синтетические: зашумленный \sin .
- Данные акселерометра.
- Данные активности мозга во время прослушивания звуковой дорожки.

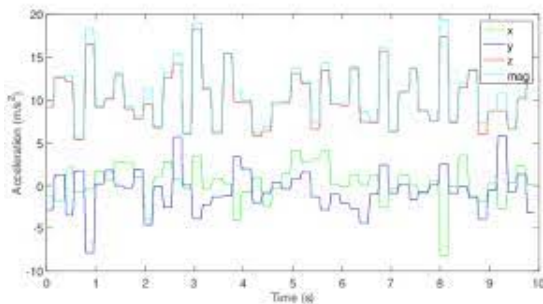


Рис.: Показания акселерометра во время подъема по лестнице.

Планируемые результаты

- Провести метрический анализ признакового пространства нейронов сети.
- Научиться выявлять ненужные нейроны.
- Научиться выделять структуру сообществ в нейронах.
- Преобразовать признаковое пространство, объединяя сильно скореллированные нейроны в новые.
- Разработать алгоритм снижения размерности признакового пространства.
- Ускорить обучение известных моделей работы с временными рядами, улучшить их качество, увеличить стабильность.