

Передбачення часових рядів за допомогою нейронних мереж

Студент БП-4 ПМ: Тхорук Ігор

Керівник: Надія Швай

Постановка задачі

Нехай часовий ряд складається з C змінних, та маємо історичні данні $X = \{X_1^t, X_2^t, \dots, X_C^t\}_{t=1}^L$, де L - довжина ряду в часі та X_i^t - значення i -тої змінної в час t . Завдання передбачення часового ряду полягає в передбаченні значень $\hat{X} = \{\hat{X}_1^t, \hat{X}_2^t, \dots, \hat{X}_C^t\}_{t=L+1}^{L+T}$ для T - майбутніх кроків.

Способи

Пряме передбачення (DMS)

- Передбачається прямо T кроків вперед
- Зазвичай краще працює якщо T велике

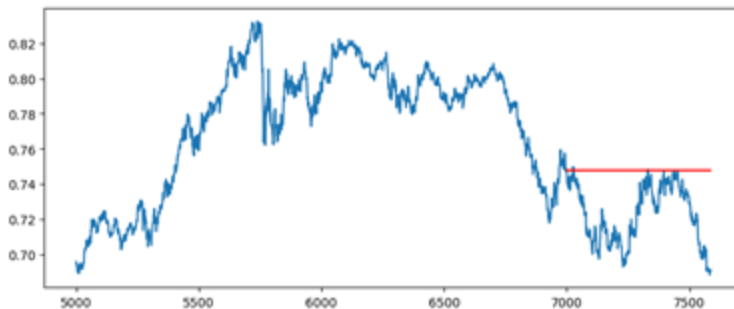
Ітеративне передбачення (IMS)

- Передбачає один крок вперед і використовує його як вхід для передбачення наступного
- Страждає від проблеми накопичення помилок
- Краще використовувати при малих T

Базовий метод

Наївний метод або рухоме середнє

Передбачення всіх майбутніх T кроків зводять до останнього відомого значення, при наївному методі або до середнього останніх L кроків при рухомому середньому.



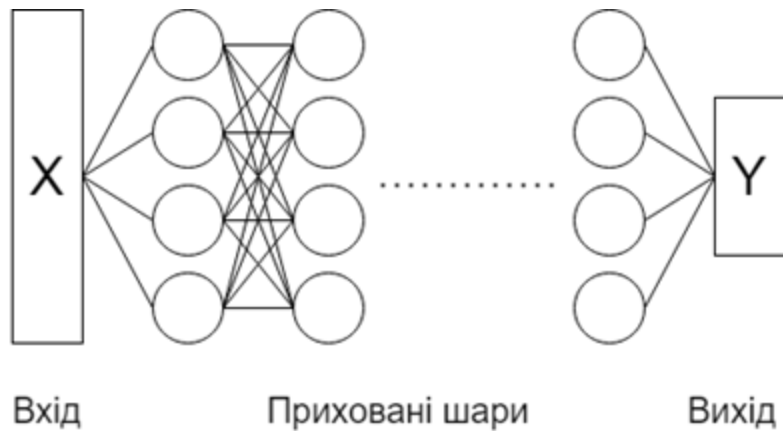
Синій - реальні значення, Червоний - передбачені значення наївним методом

Повнозв'язний шар

Формально якщо X_{n-1} вектор вихадів минулого шару (при $n = 1, X_0$ - вхідний вектор), то вектор X_n , що отриманий при застосуванні n -го шару персептрону, визначається як:

$$X_n = \sigma(W_n X_{n-1} + B_n),$$

де W_n, B_n - параметри n -го шару, що навчаються, а σ - деяка не лінійна функція активації.

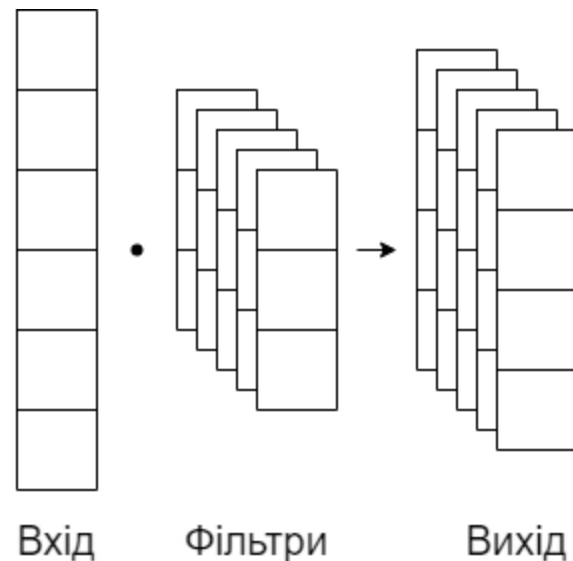


Згортковий шар

Формально, 1D згорткова операція використовується для обробки одновимірних даних, таких як часові ряди. Нехай маємо вхідний одновимірний сигнал X_i довжиною L , та фільтр (ядро) h довжиною M . Згорткова операція між X та h обчислюється наступним чином:

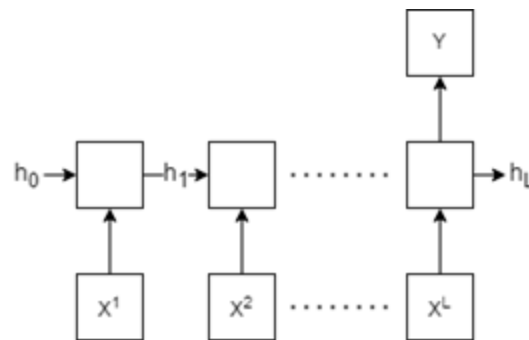
$$Y_i[t] = (X_i * h)[t] = \sum_{k=0}^{M-1} X_i[t + k - \lfloor k/2 \rfloor] h[k]$$

де Y_i - вихідний сигнал після конволюції, t - позиція у вихідному сигналі, $X[t + k - \lfloor k/2 \rfloor]$ - елемент вхідного сигналу на позиції $t + k - \lfloor k/2 \rfloor$, $h[k]$ - елемент фільтру (ядра) на позиції k .



Рекурентні мережі

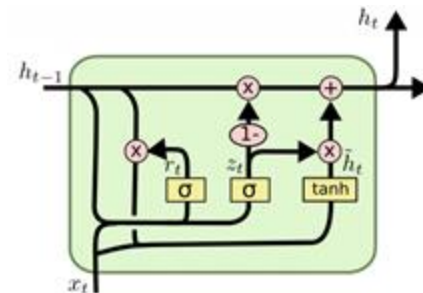
Рекурентні нейронні мережі (RNN) обробляють кожне значення ряду по чергові і використовують цю інформацію при обробці наступних значень. Це дозволяє приймати на вхід ряди будь-якої довжини.



GRU

$$\begin{aligned}Z^t &= \sigma_g(W_z X^t + U_z H^{t-1} + B_z), \\R^t &= \sigma_g(W_r X^t + U_r H^{t-1} + B_r), \\ \hat{H}^t &= \sigma_h(W_h X^t + U_h(R^t \odot H^{t-1}) + B_h), \\ H^t &= (1 - Z^t) \odot H^{t-1} + Z^t \odot \hat{H}^t,\end{aligned}$$

де X^t -вхідний вектор, H^{t-1} - внутрішній стан минулого кроку, Z^t - вектор воріт обновлення, R^t - вектор воріт відкидання, \hat{H}^t - вектор кандидат, H^t - внутрішній стан кроку, W, U, B - параметри

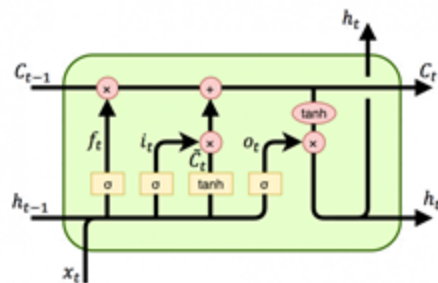


GRU

LSTM

$$\begin{aligned}F^t &= \sigma_g(W_f X^t + U_f H^{t-1} + B_f), \\I^t &= \sigma_g(W_i X^t + U_i H^{t-1} + B_i), \\O^t &= \sigma_g(W_o X^t + U_o H^{t-1} + B_o), \\ \hat{C}^t &= \sigma_h(W_c V^t + U_c H^{t-1} + B_c), \\C^t &= F^t \odot C^{t-1} + I^t \odot \hat{C}^t, \\H^t &= O^t \odot \sigma_h(C^t),\end{aligned}$$

де X^t - вхідний вектор, H^{t-1} - внутрішній стан минулого кроку, σ_g - сигмоїда, σ_h - функція \tanh , F^t - вектор воріт забування, I^t - вектор воріт оновлення, O^t - вектор воріт виходу, \hat{C}^t - вектор кандидата довгострокової пам'яті, C^t - вектор довгострокової пам'яті, H^t - внутрішній стан кроку або вихід, W, U, B - параметри



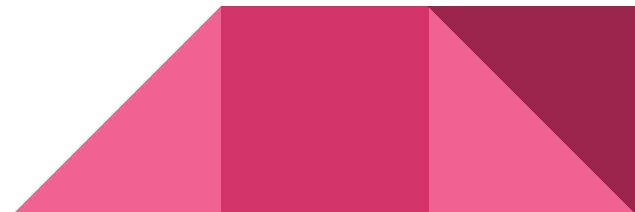
LSTM

Функція втрат

Mean Square error (MSE):

$$L(f, X, \hat{X}) = \frac{1}{T * C} \sum_{t=L}^{L+T} \sum_{i=0}^C (\hat{X}_i^t - f(X_i)^t)^2$$

де f - функція для передбачення



Дані

exchange rate - щоденні курси валют восьми іноземних країн.

solar AL - виробництво сонячної енергії за 2006 рік за кожні 10 хвилин зі 137 фотоелектричних установок штату Алабама, США.

traffic - рівень заповненості доріг (від 0 до 1) виміряний датчиками на автострадах у районі затоки Сан-Франциско, США.



Моделі

LinearNN - модель, що складається лише з двох повнозв'язних шарів

CNN_LinearNN - модель, що складається зі згорткового і двох повнозв'язних шарів

GRU - модель, що складається з GRU та повнозв'язного шарів

CNN_GRU - модель, що складається зі згорткового, GRU та повнозв'язного шарів

LSTM - модель, що складається з LSTM та повнозв'язного шарів

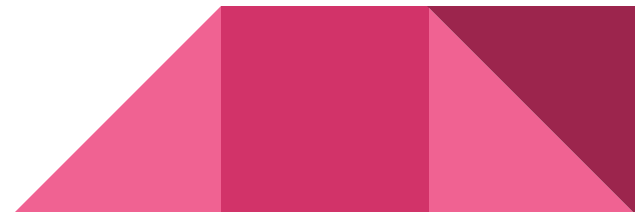
CNN_LSTM - модель, що складається зі згорткового, LSTM та повнозв'язного шарів

Результати

models	method	exchange rate				solar AL				traffic			
		1	4	16	64	1	4	16	64	1	4	16	64
LinearNN	DMS	0.009	0.011	0.017	0.044	0.012	0.035	0.118	0.217	0.242	0.357	0.415	0.419
	IMS		0.011	0.014	0.022		0.034	0.082	0.143		0.357	0.389	0.430
CNN LineraNN	DMS	0.019	0.011	0.024	0.101	0.008	0.031	0.108	0.189	0.191	0.309	0.387	0.403
	IMS		0.025	0.040	0.081		0.034	0.120	inf		0.342	5.589	1.382
GRU	DMS	0.006	0.009	0.018	0.064	0.007	0.028	0.126	0.402	0.181	0.311	0.397	0.523
	IMS		0.009	0.018	0.044		0.032	0.135	1.258		0.331	0.468	0.975
CNN GRU	DMS	0.012	0.012	0.032	0.075	0.007	0.031	0.140	0.384	0.167	0.247	0.323	0.387
	IMS		0.015	0.024	0.059		0.032	0.122	0.488		0.282	0.368	0.718
LSTM	DMS	0.008	0.012	0.021	0.077	0.007	0.028	0.141	0.396	0.169	0.284	0.359	0.461
	IMS		0.011	0.019	0.043		0.031	0.115	0.446		0.310	0.427	0.562
CNN LSTM	DMS	0.008	0.020	0.064	0.129	0.007	0.029	0.146	0.407	0.169	0.254	0.350	0.448
	IMS		0.014	0.031	0.131		0.041	0.129	0.420		0.277	0.367	0.782
Baseline	-	0.003	0.005	0.012	0.034	0.010	0.044	0.280	0.791	0.383	1.107	1.154	1.144

Висновки

З експериментів проведених в цій роботі стає зрозуміло, що жодна з моделей не є універсальним рішенням для передбачення будь-яких часових рядів. Точність моделі сильно залежить від природи використаних даних.



Дякую за увагу