

UNIwersytet Ekonomiczny w Krakowie
Instytut Metod Ilościowych w Naukach Społecznych



Jan Ślusarek

Nr. indeksu: 215302

**Porównanie efektywności metod
stochastycznych i algorytmów uczenia
maszynowego w optymalizacji portfeli
inwestycyjnych**

Praca magisterska

Pod kierunkiem
prof. UEK dr hab. Macieja Kostrzewskiego

Słowa kluczowe:

*Oświadczam, świadomy odpowiedzialności karnej za poświadczenie nie-
prawdy, że niniejszą pracę dyplomową wykonałem osobiście i samodzielnie
oraz nie korzystałem ze źródeł innych niż wymienione w pracy.*

.....

Spis treści

Wstęp	5
1. Matematyczne podwaliny budowy portfela	6
1.1. Modelowanie ciągłych procesów stochastycznych w finansach.....	6
1.2. Modelowanie dyskretnych procesów stochastycznych w finansach.....	17
1.3. Sztuczne sieci neuronowe oraz ich zastosowanie w finansach ilościowych	28
1.4. Sterowanie stochastyczne.....	36
1.5. Arbitraż statystyczny	36
1.6. Teoria optymalizacji portfela	36
2. Empiryczna budowa portfela	38
2.1. Założenia i strategia inwestycyjna	39
2.2. Podejście stochastyczne	39
2.2.1. Dobór aktywów i wag portfela.....	39
2.2.2. Modelowanie równań momentów	39
2.2.3. Analiza predykcyjna	39
2.3. Podejście algorytmiczne.....	39
2.3.1. Metody heurystyczne w doborze portfela	39
2.3.2. Algorytmy ewolucyjne w optymalizacji wag aktywów	39
2.3.3. Zastosowanie sieci neuronowych do identyfikacji wzorców rynkowych	39
2.3.4. Zastosowanie analizy skupień do segmentacji rynku	39
2.3.5. Algorytmy do oceny i zarządzania ryzykiem portfela.....	39
2.3.6. Automatyzacja procesów inwestycyjnych za pomocą algorytmów	39
2.3.7. Integracja zewnętrznych danych do analizy algorytmicznej.....	39
2.3.8. Backtesting i walidacja strategii algorytmicznych	39
3. Wyniki oraz podsumowanie	40
3.1. Wynik finansowy portfeli	40

3.2. Odporność systemów w czasie	40
3.3. Podsumowanie	40
Spis tabel.....	41
Spis rysunków.....	42
Spis kodów źródłowych	43

Wstep

Rozdział 1

Matematyczne podwaliny budowy portfela

1.1. Modelowanie ciągłych procesów stochastycznych w finansach

Koncepcja modelowania danych czasowych jest jedną z klasycznych aplikacji matematyki w naukach empirycznych. Za bardzo umowny początek tych metod, uznać możemy (nie umniejszając Leibnitzowi) *Philosophiae naturalis principia mathematica* autorstwa angielskiego fizyka Isaaca Newtona¹. Newton w swych pracach ustanowił podstawy mechaniki klasycznej, wprowadzając prawa ruchu oraz prawo powszechnej grawitacji. Te fundamentalne prawa miały ogromny wpływ nie tylko na rozwój nauki, ale i na praktyczne zastosowania w inżynierii. Skuteczność teorii Newtona dały podstawy filozofii determinizmu, która głosiła że możliwe jest precyzyjne określenie kauzalności każdego znanego nam stanu i procesu. Determinizm jako paradygmat odcisnął olbrzymi ślad w filozofii oświeceniowej, a jego implikacje były szczególnie widoczne w nukach humanistycznych, które błędnie zaczęły implementować metody mechanistyczne do złożonych procesów rzeczywistości². Załamanie się tej koncepcji przychodzi na przełomie XIX i XX wieku za sprawą francuskiego naukowca Henri’ego Poincaré wraz z jego słynnym problemem n ciał³. Problem ten brzmi następująco: Jak będzie się poruszać n ciał, które oddziałują na siebie wzajemnie siłami grawitacji? W przypadku kiedy jest jedno ciało to po prostu znajduje

¹M. Waszczyk, E. Szczerbicki, *Metodologiczne Aspekty Opisowego Modelowania w Naukach Ekonomicznych*, Zeszyty Naukowe Politechniki Gdańskiej, nr. 30, s.17

²A. Staruszkiewicz, *Izaak Newton a Oświecenia Francuskie*, Foton, nr. 107, s. 33-46

³M. Heller, *Rewolucja probabilistyczna*, *Studies in History and Philosophy of Modern Physics*, nr. 38, s. 204-218

się w spoczynku. Kiedy dane są dwa ciała rozwiązaniem jest trajektoria eliptyczna lub paraboliczna, jest to wtedy tzw. Prawo Keplera. Jednakże kiedy liczba ciał wynosi 3, nie istnieje rozwiązanie tego problemu w formie analitycznej. Wynika to z tego że małe zmiany w warunkach początkowych mogą prowadzić do dramatycznie różnych trajektorii ruchu. Obserwacja ta dała początek w formułowaniu procesów stochastycznych które jako jeden z pierwszych sformułował uczeń Poincaré'go Louis Bachelier, który użył ich w teorii spekulacji⁴. Procesy o naturze stochastycznej możemy rozumieć jako ewolucje systemu w czasie rządzonego przez prawa prawdopodobieństwa, a więc struktura tych procesów jest odmienna od procesów deterministycznych. Kardynalna różnica filozoficzna polega na tym że o ile proces deterministyczny jesteśmy w stanie prognozować bez błędu, gdyż znamy wszystkie warunki początkowe. O tyle w przypadku procesów stochastycznych tych warunków nie znamy, a więc prognozujemy zawsze z ryzykiem błędu. Niech (Ω, \mathcal{F}, P) będzie przestrzenią probabilistyczną, a T zbiorem indeksów reprezentującym czas. Proces stochastyczny $(X_t)_{t \in T}$ jest zdefiniowany jako⁵:

$$\forall t \in T, \quad X_t : \Omega \rightarrow \mathbb{R}^n, \quad (1.1)$$

gdzie:

- Ω jest przestrzenią zdarzeń,
- \mathcal{F} jest σ -ciałem na Ω ,
- P jest miarą prawdopodobieństwa na (Ω, \mathcal{F}) ,
- T jest zbiorem indeksów, zwykle reprezentującym czas (np. T może być przedziałem czasowym lub zbiorem dyskretnym momentów czasowych),
- t jest elementem zbioru T i reprezentuje konkretny moment w czasie,
- $X_t(\omega)$ jest wartością procesu w momencie t dla zdarzenia $\omega \in \Omega$,
- n jest wymiarem przestrzeni stanów procesu.

Obrazując to na przykładzie tematycznym, rozważmy cenę akcji pewnej firmy, która zmienia się każdego dnia handlowego. Niech (Ω, \mathcal{F}, P) będzie przestrzenią probabilistyczną, gdzie:

- Ω reprezentuje wszystkie możliwe ścieżki zmian cen akcji w czasie,

⁴L. Bachelier, *Théorie de la spéculation*, Annales scientifiques de l'É.N.S, seria 17, tom 3, s. 21-86

⁵M. Pitera, *Procesy Stochastyczne -notatki z wykładu-* (2022/2023), Kraków 2023, s. 7

- \mathcal{F} jest σ -ciałem zawierającym zdarzenia związane ze zmianami cen,
- P jest miarą prawdopodobieństwa, która przypisuje szanse różnym zdarzeniom rynkowym.

Niech T oznacza zbiór indeksów reprezentujących dni handlowe na giełdzie, a X_t cenę akcji w dniu t . Wtedy proces stochastyczny $(X_t)_{t \in T}$ jest zdefiniowany jako:

Twierdzenie 1.

$$\forall t \in T, \quad X_t : \Omega \rightarrow \mathbb{R}, \quad (1.2)$$

gdzie $X_t(\omega)$ to cena akcji w dniu t dla danej trajektorii $\omega \in \Omega$. W tym modelu, cena akcji X_t każdego dnia jest wartością numeryczną, która waha się losowo w zależności od różnych czynników rynkowych. Chociaż ogólne trendy rynkowe mogą być znane, codzienne fluktuacje cen są nieprzewidywalne i podlegają losowości.

Procesy stochastyczne możemy zaklasyfikować ze względu na zbiór indeksów T . Istnieją bowiem dwa przypadki kiedy zakładamy że: $t \in \mathbb{R}$ albo $t \in \mathbb{N}$. Prymat historyczny mają tutaj procesy czasu ciągłego, a konkretnie proces Wienera-Bacheliera⁶. Pojawił się niezależnie od siebie w pracach badaczy z zakresu finansów oraz fizyki teoretycznej. Wspomniany przeze mnie na poprzedniej stronie, uczeń Poincaré'go w 1900 roku opublikował go w pracy poświęconej analizie cen na giełdzie paryskiej. Niestety jego praca niesłusznie nie została w tych czasach doceniona. Świadczy o tym fakt że za odkrywce tego procesu uznawany jest Nobert Wiener, który sformalizował ten zapis ponad 20 lat później. Groteski dodaje fakt że, podzielił w pewnym sensie los swojego mistrza, który pomimo bycia prekursorem teorii względności nie został dostatecznie doceniony przez ówczesne środowisko intelektualne. Niezależnie od tych badaczy proces ten sformułował również Marian Smoluchowski oraz Albert Einstein w pracach nad modelowaniem ruchów Browna. Proces Wienera-Bachaliera $W(t)$, dla $t \geq 0$, jest zdefiniowany jako⁷:

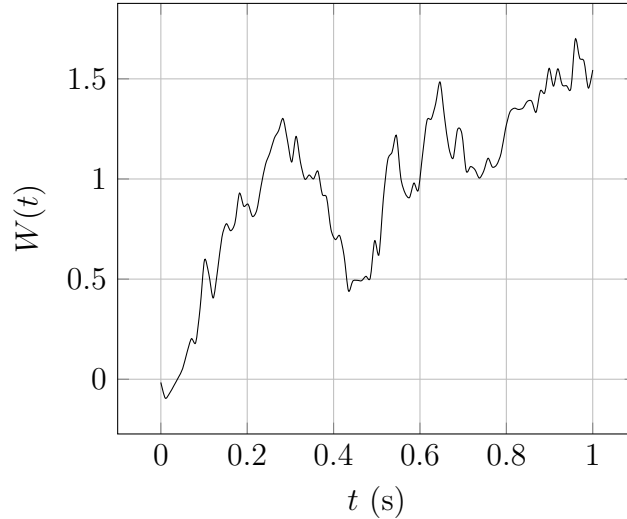
$$W(0) = 0, \quad \text{a dla } t > 0, \quad W(t) \text{ jest ciągły i posiada niezależne przyrosty.} \quad (1.3)$$

Dla $s < t$, przyrosty procesu Wienera-Bachaliera są normalnie rozłożone z wartością oczekiwaną 0 i wariancją proporcjonalną do długości przedziału czasu:

$$W(t) - W(s) \sim \mathcal{N}(0, t - s). \quad (1.4)$$

⁶J. Zabczyk, *Ważniejsze wydarzenia w teorii procesów stochastycznych*, Roczniki Polskiego Towarzystwa Matematycznego, Seria II: Wiadomości Matematyczne XL (2004), s. 78

⁷J. Jakubowski, R. Sztencel, *Wstęp do teorii prawdopodobieństwa*, Warszawa 2003, s. 304-311



Wykres 1.1: Proces Wienera-Bacheliera, opracowanie własne

- $W(t)$ oznacza wartość procesu Wienera-Bacheliera w chwili t .
- $W(s)$, gdzie s jest dowolnym momentem czasowym mniejszym niż t , reprezentuje wartość procesu Wienera w chwili s .
- t i s są zmiennymi reprezentującymi czas, z zakresu $t \geq 0$ i $s < t$.
- $\mathcal{N}(0, t - s)$ oznacza rozkład normalny z wartością oczekiwaną 0 i wariancją $t - s$.

Proces Wienera-Bacheliera charakteryzuje się ciągłością trajektorii oraz niezależnością i stacjonarnością przyrostów. Alternatywnie proces ten można zapisać jeszcze w postaci stochastycznego równania różniczkowego⁸. Rozważmy przyrost $\Delta Z(t) = Z(t) - Z(t - \Delta t)$ w krótkim przedziale czasu Δt . Z uwagi na właściwości procesu Wienera, $\Delta Z(t)$ jest zmienną losową o rozkładzie normalnym $N(0, \Delta t)$. Możemy przedstawić $\Delta Z(t)$ za pomocą zmiennej losowej ϵ o standardowym rozkładzie normalnym $N(0, 1)$, skalując ją przez $\sqrt{\Delta t}$, co prowadzi do wyrażenia $\Delta Z(t) = \epsilon \sqrt{\Delta t}$.

W kontekście stochastycznego równania różniczkowego, gdy Δt dąży do zera, stosujemy notację różniczkową:

$$dW_t = \epsilon_t \sqrt{dt} \quad (1.5)$$

gdzie ϵ_t jest standardowym białym szumem (szumem gaussowskim z niezależnymi, identycznie rozłożonymi przyrostami), a dt reprezentuje nieskończenie mały przyrost czasu.

Proces ten jest pierwszym tak ważnym przykładem modelowania stochastycznego które miało zastosowanie praktyczne w matematyce finansowej. Dalszy rozwój tej dziedziny matematyki rozwinęło teorię Analizy Stochastycznej oraz ich implementacji w finansach ilościowych. Przełomowe okazały się prace w latach 40 XX wieku Wolfganga Doeblin'a oraz Kiyosiego Itô nad stochastycznym rachunkiem różniczkowym⁹. Szczególnie ważna jest tutaj konstrukcja całki stochastycznej, która jest bardziej skomplikowana niż w przypadku klasycznych całek Riemanna lub Lebesgue'a. Zwykle wykorzystuje się do tego aproksymację sumą Itô, gdzie interwały całkowania są dzielone, a funkcja podcałkowa jest ewaluowana w sposób dyskretny¹⁰. Niech $(X_t)_{t \geq 0}$ będzie procesem stochastycznym na przestrzeni probabilistycznej $(\Omega, \mathcal{F}, \mathbb{P})$ i niech \mathcal{F}_t oznacza filtrację związaną z X_t . Całka stochastyczna względem procesu stochastycznego $(X_t)_{t \geq 0}$ jest zdefiniowana jako:

$$\int_0^T H_t dX_t = \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} H_{t_i} (X_{t_{i+1}} - X_{t_i}), \quad (1.6)$$

gdzie:

- H_t jest procesem stochastycznym adaptowanym do filtracji \mathcal{F}_t ,
- $(X_t)_{t \geq 0}$ jest semimartingale, np. procesem Ito lub ruchem Browna,
- \mathcal{F}_t jest filtracją związaną z X_t (ciąg rosnących σ -ciał),
- $0 = t_0 < t_1 < \dots < t_n = T$ jest podziałem przedziału $[0, T]$ takim, że $\max_i (t_{i+1} - t_i) \rightarrow 0$ jako $n \rightarrow \infty$,
- T reprezentuje górny limit całkowania.

Kolejnym bardzo ważnym narzędziem w rachunku stochastycznym jest Lemat Itô¹¹. Pozwala on na różniczkowanie złożonych i skomplikowanych funkcji które zawierają w sobie procesy stochastyczne. Niech X_t będzie procesem Itô, który jest zdefiniowany jako

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s, \quad (1.7)$$

gdzie μ_t i σ_t są procesami adaptowanymi do filtracji \mathcal{F}_t , a W_t jest standardowym ruchem Browna. Dla funkcji gładkiej $f : \mathbb{R}^+ \times \mathbb{R} \rightarrow \mathbb{R}$, która jest dwukrotnie różniczkowalna względem x i raz względem t , lemat Itô mówi, że

$$df(t, X_t) = \left(\frac{\partial f}{\partial t} + \mu_t \frac{\partial f}{\partial x} + \frac{1}{2} \sigma_t^2 \frac{\partial^2 f}{\partial x^2} \right) dt + \sigma_t \frac{\partial f}{\partial x} dW_t. \quad (1.8)$$

⁹H. Föllmer, *On Kiyosi Itô's Work and its Impact*, Humboldt-Universität zu Berlin, s. 1-7

¹⁰S.E. Shreve, *Stochastic Calculus for Finance II Continuous-Time Models*, Springer, Nowy Jork 2004 s. 125-128

¹¹Ibidem, s. 137-151

- **Czasowa zmiana funkcji** $\frac{\partial f}{\partial t}$: Ten termin reprezentuje zmianę funkcji f w czasie, niezależnie od ruchu procesu stochastycznego.
- **Dryf procesu** $\mu_t \frac{\partial f}{\partial x}$: Jest to składnik dryfu procesu stochastycznego. Opisuje, jak średnia wartość funkcji f zmienia się z powodu deterministycznego trendu w procesie X_t .
- **Zmienność** $\frac{1}{2} \sigma_t^2 \frac{\partial^2 f}{\partial x^2}$: Ten dodatkowy termin wynika z kwadratowej zmienności procesu X_t . Jest unikatowy dla rachunku stochastycznego i reprezentuje wpływ zmienności (ryzyka) na funkcję f .
- **Losowe wahania** $\sigma_t \frac{\partial f}{\partial x} dW_t$: Ostatni termin jest związany ze stochastycznym ruchem Browna W_t . Reprezentuje on wpływ losowych wahań procesu X_t na funkcję f .

Lemat Itô wykorzystywany jest do rozwiązywania stochastycznych równań różniczkowych, które mają zastosowania w modelowaniu zmienności instrumentów finansowych na giełdzie¹². Szczególnie jeśli charakter tych instrumentów podatny jest na bardzo częste fluktuacje (zaawansowane instrumenty pochodne). Pozwala to na precyzyjne modelowanie finansowe co jest kluczowe dla zarządzania ryzykiem, hedgingu, spekulacji, i innych operacji inwestycyjnych na rynkach finansowych.

Założmy, że dynamika ceny akcji S_t jest opisana przez następujące równanie różniczkowe stochastyczne (SDE) w modelu Itô:

$$dS_t = \mu S_t dt + \sigma S_t dW_t, \quad (1.9)$$

gdzie:

- μ reprezentuje stałą stopę wzrostu (dryf) ceny akcji,
- σ jest stałą stopą zmienności ceny akcji,
- W_t jest procesem Wienera (ruchem Browna).

Rozważmy prosty przypadek, gdzie chcemy obliczyć oczekiwaną cenę akcji S_t po upływie czasu T . Zakładamy, że cena początkowa S_0 jest znana. Ponieważ równanie ma postać stochastyczną, używamy lematu Itô do jego rozwiązania. Otrzymujemy:

¹²A.G. Malliaris, *Itô Calculus in Financial Decision Making*, Society for Industrial and Applied Mathematics Review, Tom 25 nr. 4, s. 481-485

$$\log(S_T) = \log(S_0) + \left(\mu - \frac{1}{2}\sigma^2\right)T + \sigma W_T. \quad (1.10)$$

Wykorzystując własności procesu Wienera, możemy wyznaczyć oczekiwaną wartość S_T . Oczekiwana wartość logarytmu ceny jest równa:

$$\mathbb{E}[\log(S_T)] = \log(S_0) + \left(\mu - \frac{1}{2}\sigma^2\right)T. \quad (1.11)$$

Stąd oczekiwana cena akcji w chwili T wynosi:

$$\mathbb{E}[S_T] = S_0 e^{(\mu - \frac{1}{2}\sigma^2)T}. \quad (1.12)$$

Przeanalizujmy teraz n akcji, których ceny są reprezentowane przez procesy stochastyczne $S_t^1, S_t^2, \dots, S_t^n$. Dynamika każdej akcji jest opisana przez równanie różniczkowe stochastyczne (SDE) w modelu Itô:

$$dS_t^i = \mu_i S_t^i dt + \sum_{j=1}^n \sigma_{ij} S_t^i dW_t^j, \quad i = 1, 2, \dots, n, \quad (1.13)$$

gdzie:

- μ_i reprezentuje stałą stopę wzrostu (dryf) i -tej akcji,
- σ_{ij} jest elementem macierzy zmienności, opisującej współzależności między różnymi akcjami,
- W_t^j są procesami Wienera, które mogą być skorelowane między sobą.

Obliczmy oczekiwane wartości cen tych akcji w przyszłości. Zakładając, że cena początkowa S_0^i każdej akcji jest znana, możemy użyć wektorowego podejścia do obliczeń.

Oczekiwana wartość logarytmu ceny i -tej akcji w chwili T wynosi:

$$\mathbb{E}[\log(S_T^i)] = \log(S_0^i) + \left(\mu_i - \frac{1}{2} \sum_{j=1}^n \sigma_{ij}^2\right)T. \quad (1.14)$$

Oczekiwana cena i -tej akcji w chwili T wynosi:

$$\mathbb{E}[S_T^i] = S_0^i e^{\left(\mu_i - \frac{1}{2} \sum_{j=1}^n \sigma_{ij}^2\right)T}. \quad (1.15)$$

Powyższy model pozwala na analizie portfela akcji, które uwzględnia indywidualne ryzyko danego aktywa jak i również ryzyko związane ze współzależnościami.

Możliwość szerokiego zastosowania SDE do modelowania ruchów giełdowych, poprzez narzędzia analizy stochastycznej, pozwoliło na implementacje różnorodnych równań. Poniżej wymienię najważniejsze z nich.^{13 14 15 16 17 18 19}

Proces Geometrycznego Ruchu Browna

$$dS_t = \mu S_t dt + \sigma S_t dW_t \quad (1.16)$$

gdzie:

- S_t - cena aktywa w czasie t ,
- μ - stały współczynnik wzrostu,
- σ - zmienność ceny aktywa,
- W_t - proces Wienera.

Geometryczne ruchy Browna są najbardziej podstawowym ciągłym opisem cen akcji lub innych instrumentów finansowych. W tym modelu ceny ulegają ciągłym, losowym fluktuacjom a parametr μ określa tendencję wzrostową albo spadkową.

Równanie Mertona

$$dS_t = \mu S_t dt + \sigma S_t dW_t + dJ_t \quad (1.17)$$

gdzie:

- S_t - cena aktywa w czasie t ,
- μ - stały współczynnik wzrostu,
- σ - zmienność ceny aktywa,

¹³B. Øksendal, *Stochastic Differential Equations An Introduction with Applications*, Nowy Jork 2002, Springer, s. 612-614

¹⁴R. Merton, *Option Pricing When Underlying Stock Returns Are Discontinuous*, Journal of Financial Economics, nr. 3, s. 125-144

¹⁵J.C. Cox, J.E. Ingersoll, S.A. Ross, *A Theory of the Term Structure of Interest Rates*, Econometrica, Tom 63 nr. 2, s. 390

¹⁶E.M. Stein, J.C. Stein, *Stock Price Distributions with Stochastic Volatility: An Analytic Approach*, The Review of Financial Studies, Tom 4 nr. 4, s. 727-752

¹⁷S.L. Heston, *A closed-form solution for options with stochastic volatility with applications to bond and currency options*, The Review of Financial Studies, Tom 6 nr. 2, s. 327-343

¹⁸Y. Chang, Y. Wang, S. Zhang, *Option Pricing under Double Heston Model with Approximative Fractional Stochastic Volatility*, Mathematical Problems in Engineering, Tom 2021, s. 1-12

¹⁹D.S. Bates, *Jumps and Stochastic volatility: Exchange Rate Processes Implicity in Deutsche Mark Options*, The Review of Financial Studies, Tom 9 nr. 1, s. 69-107

- W_t - proces Wienera,
- J_t - skok, reprezentujący nagłe zmiany ceny.

Jest to w istocie równanie Geometrycznego Ruchu Browna ale z dodatkowym parametrem J_t , który uwzględnia nagłe skoki cenowe.

Proces Ornsteina-Uhlenbecka

$$dX_t = \theta(\mu - X_t) dt + \sigma dW_t \quad (1.18)$$

gdzie:

- X_t - proces stochastyczny w czasie t ,
- θ - szybkość powrotu do średniej,
- μ - średnia długoterminowa,
- σ - zmienność procesu,
- W_t - proces Wienera.

Proces ten jest szczególnie użyteczny jeśli dany aktyw cechuje się powrotem do średniej. Wykorzystywany w modelowaniu stóp procentowych, tempa inflacji, kursów walutowych czy w algorytmicznych strategiach handlowych.

Model Coxa-Ingersolla-Rossa

$$dr_t = \kappa(\theta - r_t) dt + \sigma\sqrt{r_t} dW_t \quad (1.19)$$

gdzie:

- r_t reprezentuje stopę procentową w czasie t ,
 - κ to współczynnik powrotu do średniej, określający szybkość, z jaką stopa procentowa dąży do swojej średniej długoterminowej θ ,
 - θ jest długoterminową średnią stopy procentowej,
 - σ to zmienność stopy procentowej,
 - dW_t jest przyrostem procesu Wienera.
-

Jest to rozszerzenie procesu Ornsteina-Uhlenbecka o dodatkowe założenia, gdzie dodatkowy parametr stopy procentowej pod pierwiastkiem nie pozwala przyjmować stopie procentowej wartości ujemne. Jest to szczególnie przydatne w modelowaniu krótkoterminowej stopy procentowej.

Model Hestona

$$dS_t = \mu S_t dt + \sqrt{v_t} S_t dW_t^S \quad (1.20)$$

$$dv_t = \kappa(\theta - v_t) dt + \sigma \sqrt{v_t} dW_t^v \quad (1.21)$$

gdzie:

- S_t - cena aktywa w czasie t ,
- μ - oczekiwany zwrot (dryf) aktywa,
- v_t - zmienność aktywa w czasie t ,
- κ - szybkość powrotu zmienności do długoterminowej średniej θ ,
- θ - długoterminowa średnia zmienność,
- σ - zmienność zmienności, znana również jako "vol of vol",
- W_t^S i W_t^v - dwa procesy Wienera, które mogą być skorelowane.

W modelu Hestona szczególnie użyteczną funkcją jest możliwość modelowania nie tylko ceny aktywa ale i również jego stochastycznej zmienności. Jest on szczególnie przydatny kiedy sama zmienność podlega nieliniowym fluktuacjom

Podwójny Model Hestona

$$dS_t = \mu S_t dt + \sqrt{v_{1,t} + v_{2,t}} S_t dW_t^S \quad (1.22)$$

$$dv_{1,t} = \kappa_1(\theta_1 - v_{1,t}) dt + \sigma_1 \sqrt{v_{1,t}} dW_t^{v1} \quad (1.23)$$

$$dv_{2,t} = \kappa_2(\theta_2 - v_{2,t}) dt + \sigma_2 \sqrt{v_{2,t}} dW_t^{v2} \quad (1.24)$$

gdzie:

- S_t - cena aktywa w czasie t ,
- μ - oczekiwany zwrot (dryf) aktywa,
- $v_{1,t}$ i $v_{2,t}$ - dwie składowe zmienności aktywa w czasie t ,
- $\kappa_1, \theta_1, \sigma_1$ oraz $\kappa_2, \theta_2, \sigma_2$ - parametry modelu dla obu składowych zmienności,

- dW_t^S, dW_t^{v1} i dW_t^{v2} - procesy Wienera dla ceny aktywa i obu składowych zmienności.

Powyższy model jest uogólnieniem modelu Hestona na dwa składowe stochastycznej zmienności. Dzięki temu zabiegowi, możemy uchwycić coraz to bardziej nieliniowe fluktuacje w zmienności cen.

Model SVJ

$$dS_t = \mu S_t dt + \sqrt{v_t} S_t dW_t^S + J_t S_t dN_t \quad (1.25)$$

$$dv_t = \kappa(\theta - v_t) dt + \sigma \sqrt{v_t} dW_t^v \quad (1.26)$$

gdzie:

- S_t - cena aktywa w czasie t ,
- μ - oczekiwany zwrot (dryf) aktywa,
- v_t - zmienność aktywa w czasie t (modelowana stochastycznie),
- J_t - wielkość skoku ceny aktywa,
- dN_t - proces Poissona reprezentujący liczbę i czas wystąpienia skoków,
- κ - szybkość powrotu zmienności do długoterminowej średniej θ ,
- θ - długoterminowa średnia zmienność,
- σ - zmienność zmienności
- W_t^S i W_t^v - dwa procesy Wienera, które mogą być skorelowane.

Model SVJ (Stochastic Volatility with Jumps) jest w pewnym sensie synergia między równaniem Mertona a modelem Hestona. Jest on szczególnie przydatny w modelowaniu cen które charakteryzują się nieliniowymi zmiennościami wraz z nagłymi spadkami lub wzrostami cen.

Postrzeganie procesu stochastycznego jako ciągłego jest zazwyczaj domeną rynku instrumentów pochodnych, aczkolwiek nie jest to uniwersalna reguła. Wiąże się ona z faktem, że w popularnych modelach na wycenę opcji na przykład wzór Blacka-Scholesa-Mertona uznaje że zmiana ceny instrumentu bazowego ma postać stochastycznego równania różniczkowego. Przyjęło się wśród matematyków finansowych za pośrednictwem teorii Bacheliera *Théorie de la spéculation*²⁰, modelowanie ruchów giełdowych jako procesów

²⁰L. Bachelier, *Théorie de la spéculation*, Annales scientifiques de l'É.N.S, seria 17, tom 3, s. 21-86

ciągłych, co później niezależnie kontynowali²¹: Myron Scholes, Fisher Black oraz Robert Merton. Złożone równania SDE mogą bardzo dobrze odzwierciedlać bardzo szybkie nieliniowe zmiany w czasie, ponadto z zasad całkowania stochastycznego cechują się one większą czytelnością obliczeniową niż jakbyśmy chcieli postrzegać te procesy jako dyskretne. Niemniej jednak warto tutaj zaznaczyć że te procesy są ciągłe ale tylko w teorii, bo w praktyce są one dyskretne, w tym sensie, że i tak równania SDE dyskretyzujemy i kalibrujemy do danych empirycznych.

1.2. Modelowanie dyskretnych procesów stochastycznych w finansach

Odmiernym podejściem do modelowania zmienności cen aktywów jest postrzeganie czasu jako zbioru dyskretnego. Intuicyjnie takie podejście wydaje się bardziej odzwierciedlać pracę giełdy, bowiem zmiany wyceny aktywów na rynku finansowym są rejestrowane w sztywnych godzinach handlowych, a cena podlega prawu popytu i podaży²². Załóżmy, że mamy daną akcję na rynku, a jej cena w czasie t jest oznaczona jako P_t . Zmiana ceny akcji może być opisana w uproszczony sposób, biorąc pod uwagę popyt (D_t) i podaż (S_t) akcji:

$$\Delta P_t = f(D_t, S_t), \quad (1.27)$$

gdzie ΔP_t oznacza zmianę ceny, a funkcja f opisuje zależność między popytem a podażą. Możemy to uproszczenie wyrazić jako:

- Jeśli $D_t > S_t$, czyli popyt przewyższa podaż, cena akcji wzrasta.
- Jeśli $D_t < S_t$, czyli podaż przewyższa popyt, cena akcji spada.
- Jeśli $D_t = S_t$, cena akcji pozostaje stabilna.

Równanie to można wyrazić w bardziej formalny sposób, na przykład poprzez określenie funkcji f , która mogłaby uwzględniać różnice między popytem a podażą oraz inne czynniki rynkowe. Powyższe równania i założenia pozwolą nam lepiej zrozumieć prawa jakimi rządzi się stopa zwrotu.

²¹J. Choi, M. Kwak, C.W. Tee, Y. Wang, *A Black-Scholes user's guide to the Bachelier model*, The Journal of Futures Markets, Tom 42 nr. 5, s. 959-962

²²O. Blanchard, *Makroekonomia*, Warszawa 2016, s. 128-134

Historycznie rzecz biorąc modele dyskretne, swoje pochodzenie wywodzą z makroekonomicznej. A konkretnie z prac Christophera Simsa²³, Roberta Engle²⁴, Clive'a Granger'a²⁵, które później na płaszczyźnie finansów z sukcesem zaaplikował Tim Bollerslev²⁶. W tym przypadku interesują nas procesy które są kowariancyjnie stacjonarne (słabo stacjonarne). Dzieje się tak dlatego że dyskretne szeregi czasowe modelujemy jako funkcję swoich poprzednich wartości za pomocą modeli AR(ang. *Autoregressive*). Zdefiniujmy najpierw silną stacjonarność. Proces stochastyczny $\{X_t\}$ jest silnie stacjonarny²⁷, jeśli dla dowolnych momentów czasowych t_1, t_2, \dots, t_n i dla każdego przesunięcia czasowego τ , mamy:

$$P(X_{t_1}, X_{t_2}, \dots, X_{t_n}) = P(X_{t_1+\tau}, X_{t_2+\tau}, \dots, X_{t_n+\tau}) \quad (1.28)$$

Tak więc powyższe równanie można odczytać że, proces jest silnie stacjonarny wtedy kiedy jego wspólny rozkład prawdopodobieństwa nie zmienia się w czasie. Słaba stacjonarność (kowariancyjna stacjonarność) występuje wtedy kiedy wartość oczekiwana i wariancja jest stała w czasie, jednakże kowariancja jest funkcją różnicy między momentami czasowymi. Proces stochastyczny $\{X_t\}$ jest słabo stacjonarny, jeśli spełnia następujące warunki:

1. **Stała średnia:** Średnia procesu jest stała w czasie. Dla każdego czasu t , średnia procesu, oznaczana jako $E[X_t]$, jest równa pewnej stałej μ , niezależnie od t . Możemy to zapisać jako:

$$E[X_t] = \mu, \quad \forall t. \quad (1.29)$$

2. **Stała wariancja:** Wariancja procesu jest stała w czasie. Oznacza to, że dla każdego t , wariancja $Var[X_t]$ jest równa pewnej stałej σ^2 , niezależnie od t . Możemy to wyrazić jako:

$$Var[X_t] = E[(X_t - \mu)^2] = \sigma^2, \quad \forall t. \quad (1.30)$$

3. **Stała kowariancja w czasie:** Dla każdej pary czasów t_1 i t_2 , kowariancja między X_{t_1} a X_{t_2} zależy tylko od różnicy czasu $\tau = |t_2 - t_1|$, a nie od konkretnych wartości t_1 i t_2 . Innymi słowy, kowariancja jest funkcją różnicy czasu τ , oznaczaną jako $\gamma(\tau)$:

$$Cov(X_{t_1}, X_{t_2}) = E[(X_{t_1} - \mu)(X_{t_2} - \mu)] = \gamma(\tau), \quad \forall t_1, t_2. \quad (1.31)$$

²³Ch. Sims, *Macroeconomics and Reality*, Econometrica tom 48 nr. 1, s. 1-48

²⁴R.F. Engle, *Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation*, Econometrica tom 50 nr. 4, s. 987-1007

²⁵R.F. Engle, C.W.J. Granger, *Co-integration and error correction: Representation, estimation and testing*, Econometrica tom 55 nr. 2, s. 251-276

²⁶T. Bollerslev, *A Conditional Heteroskedastic Time Series Model for Speculative Prices and Rates of Return*, The Review of Economics and Statistics tom 63 nr. 3, s. 542-547

²⁷P.J. Brockwell, R.A. Davis, *Time Series: Theory and Methods*, wydanie 2, Nowy Jork 2006, s. 11-14

Jako że prawie nigdy ceny aktywów finansowych na rynkach kapitałowych nie spełniają wymogu słabej (kowariancyjnej) stacjonarności. Zmuszeni jesteśmy wykonać operacje różnicowania ceny, albo co jest też bardziej popularnym podejściem pracować na stopach zwrotu²⁸. Najpopularniejsze rodzaje stóp zwrotu to²⁹:

(a) **Prosta stopa zwrotu**

$$R = \frac{P_t - P_{t-1}}{P_{t-1}} \quad (1.32)$$

gdzie:

- R - prosta stopa zwrotu,
- P_t - cena aktywa w chwili t ,
- P_{t-1} - cena aktywa w chwili $t - 1$.

(b) **Logarytmiczna stopa zwrotu**

$$R = \ln \left(\frac{P_t}{P_{t-1}} \right) \quad (1.33)$$

gdzie:

- R - logarytmiczna stopa zwrotu,
- \ln - logarytm naturalny,
- P_t - cena aktywa w chwili t ,
- P_{t-1} - cena aktywa w chwili $t - 1$.

(c) **Stopa zwrotu z dywidendami**

$$R = \frac{P_t + D - P_{t-1}}{P_{t-1}} \quad (1.34)$$

gdzie:

- R - stopa zwrotu z dywidendami,
- P_t - cena aktywa w chwili t ,
- D - dywidenda wypłacona w okresie,
- P_{t-1} - cena aktywa w chwili $t - 1$.

(d) **Stopa zwrotu skorygowana o ryzyko (np. Sharpe Ratio)**

$$R_{\text{risk-adjusted}} = \frac{R - R_f}{\sigma} \quad (1.35)$$

gdzie:

²⁸M. Doman, R. Doman, *Modelowanie zmienności i ryzyka - metody ekonometrii finansowej*, Kraków 2009, s. 20-34

- $R_{\text{risk-adjusted}}$ - stopa zwrotu skorygowana o ryzyko,
- R - stopa zwrotu aktywa,
- R_f - stopa zwrotu wolna od ryzyka,
- σ - odchylenie standardowe stopy zwrotu aktywa.

W kontekście równania 1.27 które wyprowadziłem na początku podrozdziału, kowariancyjna stacjonarność, na przykład prostych stóp zwrotu (co zazwyczaj ma miejsce na rynkach finansowych) mówi że, w danym horyzoncie czasowym na rynku występuje w miarę stała przewaga kupna względem sprzedaży albo odwrotnie. Zależy to od horyzontu czasowego w którym analizujemy dany szereg. Zdefiniujmy teraz proces autoregresyjny rzędu k , oznaczany jako $AR(k)$, jest definiowany jako³⁰:

$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_k X_{t-k} + \varepsilon_t, \quad (1.36)$$

gdzie:

- X_t jest wartością procesu w czasie t ,
- $\phi_0, \phi_1, \phi_2, \dots, \phi_k$ są parametrami procesu,
- ε_t jest białym szumem o średniej równej 0 i stałej wariancji.

Parametry $\phi_1, \phi_2, \dots, \phi_k$ reprezentują wpływ k poprzednich wartości procesu na obecną wartość X_t . Wartość ϕ_0 jest stałą, która pozwala na przesunięcie całego procesu w górę lub w dół na osi wartości. Szczególnie ważne w kontekście tego procesu jest twierdzenie Wolda³¹ które głosi że, niech $\{X_t\}$ będzie słabo stacjonarnym procesem stochastycznym. Wówczas $\{X_t\}$ można przedstawić jako sumę dwóch składowych: deterministycznej $\{D_t\}$ i stochastycznej $\{S_t\}$, gdzie:

$$X_t = D_t + S_t, \quad (1.37)$$

z zastrzeżeniem, że:

- D_t jest deterministycznym składnikiem procesu, który można przedstawić jako skończoną sumę przesunięć czasowych wartości procesu.

³⁰G. Udny Yule, *On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers*, Philosophical Transactions of the Royal Society, Tom 226, s. 267-298

³¹A. Nielsen, *Szeregi czasowe. Praktyczna analiza i predykcja z wykorzystaniem statystyki i uczenia maszynowego*, Gliwice 2020, s. 179-181

- S_t jest stochastycznym składnikiem procesu, który jest liniowym, przyczynowym i odwracalnym filtrem białego szumu $\{\varepsilon_t\}$, co oznacza, że:

$$S_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad (1.38)$$

gdzie $\psi_0 = 1$ i szereg $\sum_{j=0}^{\infty} \psi_j^2$ jest zbieżny.

Twierdzenie to podkreśla, że każdy słabo stacjonarny proces stochastyczny można przedstawić jako sumę deterministycznego składnika oraz nieskorelowanego z nim składnika stochastycznego, który jest liniową funkcją białego szumu. Konsekwencje Twierdzenia Wolda w kontekście wielomianu charakterystycznego (który determinuje czy proces jest stacjonarny czy nie) wskazuje, że każdy słabo stacjonarny proces stochastyczny $\{Y_t\}$ może być przedstawiony jako:

$$Y_t = \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j} + \nu_t, \quad (1.39)$$

gdzie $\{\varepsilon_t\}$ jest białym szumem, a $\{\nu_t\}$ reprezentuje deterministyczną część procesu.

Przy użyciu operatora opóźnienia L , równanie to można przekształcić do postaci:

$$Y_t = \Psi(L)\varepsilon_t + \nu_t, \quad (1.40)$$

gdzie $\Psi(L) = \sum_{j=0}^{\infty} \Psi_j L^j$ jest wielomianem charakterystycznym procesu.

Konsekwencje dla wielomianu charakterystycznego:

- Stabilność i odwracalność:** Aby proces był stacjonarny, wszystkie pierwiastki wielomianu $\Psi(z)$ (gdzie z jest zmienną zespoloną) muszą znajdować się poza okręgiem jednostkowym na płaszczyźnie zespolonej. To oznacza, że $|\Psi(z)| = 0$ implikuje $|z| > 1$.
- Odwracalność:** Jeżeli dla każdego $|z| \neq 1$, istnieje $\Theta(z)$ takie, że $\Psi(z)\Theta(z) = \Theta(z)\Psi(z) = I_n$, to proces jest odwracalny. Oznacza to, że istnieje odwrotny filtr $\Theta(L)$, taki że $\Theta(L)Y_t = \mu_t + \varepsilon_t$, gdzie μ_t jest składnikiem deterministycznym.
- Reprezentacja procesu:** Proces Y_t można wówczas przedstawić jako:

$$Y_t = - \sum_{i=1}^{\infty} \Theta_i Y_{t-i} + \mu_t + \varepsilon_t, \quad (1.41)$$

gdzie Θ_i są współczynnikami odwrotnego filtra.

Te konsekwencje podkreślają znaczenie wielomianu charakterystycznego w zasadnym modelowaniu szeregów czasowych. Ogólnie to można generalizować model AR i przedstawić go również w formie ARIMA. Definicja Modelu ARIMA (p, d, q): Model ARIMA (p, d, q), gdzie p oznacza rząd autoregresji, d stopień różnicowania, a q rząd średniej ruchomej, jest zdefiniowany jako³²:

$$(1 - \sum_{i=1}^p \phi_i L^i)(1 - L)^d X_t = (1 + \sum_{j=1}^q \theta_j L^j) \varepsilon_t, \quad (1.42)$$

gdzie:

- X_t jest obserwowaną wartością szeregu czasowego w czasie t ,
- ϕ_i są parametrami modelu autoregresyjnego,
- θ_j są parametrami modelu średniej ruchomej,
- ε_t jest białym szumem (zakłóceniem) o średniej równej zero i stałej wariancji,
- L jest operatorem opóźnienia (Lag operator), gdzie $L^k X_t = X_{t-k}$,
- $(1 - L)^d$ jest operatorem różnicowania stopnia d .

Model ARIMA jest więc kombinacją trzech podstawowych elementów:

- (a) **Autoregresja (AR):** Zależność między obecną wartością a wartościami z poprzednich kroków czasowych.
- (b) **Różnicowanie (I):** Różnicowanie szeregu w celu uzyskania stacjonarności.
- (c) **Średnia Ruchoma (MA):** Zależność między obecną wartością a błędami z poprzednich kroków czasowych.

Do estymacji modeli autoregresyjnych najczęściej wykorzystuje się popularną metodę MNK (Metoda Najmniejszych Kwadratów), bądź też poprzez: MNW (Metoda Największej Wiarygodności), równania Yule’a-Walkera czy metody bayesowskie.

W kontekście danych finansowych warto tutaj nadmienić że często obserwowanym zjawiskiem w stopach zwrotu jest to, kiedy funkcja autokorelacji nagle skokowo odwołuje się do przeszłości. Nawet wtedy kiedy szereg jest stacjonarny. Wtedy pomocna może być technika frakcyjnego różnicowania. Metoda ta polega na tym, że parametr d w modelu ARIMA(p, d, q) (równanie numer 1.38) jest liczbą \mathbb{R} w zakresie od -1 do 1. Wtedy model taki nosi nazwę ARFIMA (ang. *Autoregressive Fractionally Integrated Moving Average*).

³²P.J. Brockwell, R.A. Davis, *Time Series: Theory and Methods*, wydanie 2, Nowy Jork 2006, s. 274

Modelowanie stóp zwrotu nie jest wcale zadaniem łatwym, dane finansowe mają to do siebie że występuje w nich mocna i widoczna heteroskedastyczność oraz grupowanie zmienności (ang. *volatility clusters*)³³, oznacza to, że wariancja w czasie nie jest stała. Obrazując to na przykładzie, w szeregach czasowych mówimy o heteroskedastyczności, gdy wariancja błędów losowych zmienia się w czasie. Formalnie, dla szeregu czasowego $\{X_t\}$ z modelem

$$X_t = \mu_t + \varepsilon_t, \quad (1.43)$$

gdzie μ_t jest wartością oczekiwaną w czasie t , a ε_t jest błędem losowym, heteroskedastyczność występuje, gdy

$$\text{Var}(\varepsilon_t) \neq \text{const}. \quad (1.44)$$

W przypadku rynków finansowych oznacza to że istnieją różne okresy zmienności cen aktywów (raz duże a raz małe). Zjawisko to nosi nazwę efektu ARCH, w którym modeluje się również wariancję jako funkcję warunkową względem swojej własnej przeszłości. Do najpopularniejszych i najogólniejszych metod modelowania tego zjawiska wykrzysytuje się model GARCH (Generalized Autoregressive Conditional Heteroskedasticity) jest on rozszerzeniem modelu ARCH, odkrytym przez Bollersleva³⁴. Dla modelu GARCH(p, q), mamy

$$\varepsilon_t = \sigma_t z_t, \quad (1.45)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2, \quad (1.46)$$

gdzie:

- ε_t i σ_t^2 są zdefiniowane jak w modelu ARCH,
- β_j są dodatkowymi parametrami modelu,
- q określa liczbę opóźnień wariancji w modelu.

Istnieje wiele rozszerzeń tego modelu, które uwzględniają do tego różnorakie zmienne np. GJR-GARCH, jednakże warto tutaj zaznaczyć że forma z równania 1.46 musi występować w każdej modyfikacji modeli z tej klasy. Jeżeli modelujemy stopę zwrotu jako proces autoregresyjny a reszty jako proces GARCH, to możemy sobie wyobrazić

³³M. Doman, R. Doman, *Modelowanie zmienności i ryzyka - metody ekonometrii finansowej*, Kraków 2009, s. 29-30

³⁴T. Bollerslev, *Generalized Autoregressive Conditional Heteroskedasticity*, Journal of Econometrics, tom 31, nr. 3, s. 542-547

że rozkład który generuje wartości dla danego procesu stochastycznego w pewnym sensie jest ruchomy w czasie względem momentu zwykłego i centralnego. Ponadto zmienność wariancji możemy rozumieć jako estymator ryzyka (bowiem jest on funkcją odchyień od wartości oczekiwanej), co jest szczególnie przydatne w zarządzaniu ryzykiem.

Wymienione przeze mnie w tym podrozdziale metody dotyczą analizy jednowymiarowej. Oznacza to, że interesuje nas w tym przypadku tylko jedno aktyw, które staramy się ubrać w język matematyki. Jednakże z perspektywy tej pracy interesuje mnie analiza portfela aktyw, a więc należy rozszerzyć równania momentu zwykłego i centralnego ze skalaru w wektor. Na początku rozszerzymy proces jednowymiarowej autoregresji AR w proces wektorowej autoregresji VAR wraz z wszystkimi tego konsekwencjami³⁵.

$$\mathbf{Y}_t = \mathbf{c} + \mathbf{A}_1 \mathbf{Y}_{t-1} + \mathbf{A}_2 \mathbf{Y}_{t-2} + \dots + \mathbf{A}_p \mathbf{Y}_{t-p} + \mathbf{u}_t, \quad (1.47)$$

gdzie:

- \mathbf{Y}_t jest $n \times 1$ wektorem zmiennych w czasie t ,
- \mathbf{c} jest $n \times 1$ wektorem stałych (interceptów),
- $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_p$ są $n \times n$ macierzami współczynników autoregresji,
- \mathbf{u}_t jest $n \times 1$ wektorem błędów (szumów) w czasie t , zwykle zakłada się, że ma rozkład normalny z średnią zero i macierzą kowariancji Σ ,
- p jest rzędem autoregresji, określającym liczbę opóźnień w modelu.

Żeby lepiej zrozumieć sens tego modelu, przedstawmy prosty przykład. Załóżmy, że mamy trzy aktywa w naszym portfelu, a ich ceny (lub stopy zwrotu) w czasie są reprezentowane przez zmienne $Y_{1,t}, Y_{2,t}, Y_{3,t}$. Model VAR(1) dla tego portfela może być zapisany jako:

$$\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \\ Y_{3,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \\ Y_{3,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \end{bmatrix}, \quad (1.48)$$

gdzie:

- $Y_{i,t}$ jest ceną (lub stopą zwrotu) i -tego aktywa w czasie t ,
- c_i jest stałą (interceptem) dla i -tego aktywa,

³⁵H. Lutkepohl, M. Kratzig, *Applied time series econometrics*, Nowy Jork 2004, s. 88

- a_{ij} są współczynnikami autoregresji, pokazującymi, jak ceny aktywów w przeszłości wpływają na cenę i -tego aktywa,
- $u_{i,t}$ jest błędem (szumem) dla i -tego aktywa w czasie t .

W tym modelu, każde aktywo jest modelowane jako zależne nie tylko od własnych poprzednich cen, ale również od cen pozostałych aktywów w portfelu. Taki model pozwala na analizę wzajemnych zależności między aktywami i jest przydatny w prognozowaniu przyszłych cen i zwrotów aktywów w portfelu. Forma modelu dziedziczy wszystkie właściwości ARIMA (p,d,q), z tym że, średnie ruchome są tutaj wektorami. Jeżeli badamy proces VAR opisany równaniem 1.47, gdzie $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_p$ są macierzami współczynników autoregresji, a \mathbf{u}_t jest wektorem błędów. Macierz stowarzyszona³⁶ \mathbf{M} dla tego modelu VAR(p) jest macierzą o wymiarach $np \times np$ i ma postać:

$$\mathbf{M} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_{p-1} & \mathbf{A}_p \\ \mathbf{I}_n & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_n & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_n & \mathbf{0} \end{bmatrix}, \quad (1.49)$$

gdzie:

- \mathbf{I}_n jest $n \times n$ macierzą jednostkową,
- $\mathbf{0}$ jest $n \times n$ macierzą zerową.

Jeżeli moduł wartości własnej macierzy \mathbf{M} przekracza 1, oznacza to, że model VAR(p) nie jest stacjonarny. Operacja przekształcenia szeregu w stacjonarny jest w istocie ta sama jak dla modelu jednorównaniowego, z tym że operujemy na macierzach. W praktyce jak wspomniałem wcześniej, stopy zwrotu które zazwyczaj dominują w analizie empirycznej, cechują się kowariancyjną stacjonarnością.

Oprócz modelowania wartości oczekiwanej portfela, jako funkcji wektora wartości oczekiwanej za pomocą VAR(p), możemy również rozszerzyć jednorównaniowy model GARCH w jego wielowymiarowy przypadek MGARCH. Takie rozszerzenie pozwala na kompleksową analizę dynamiki zmienności i kowariancji między wieloma aktywami. Modele MGARCH są niezwykle przydatne w modelowaniu i prognozowaniu zmienności rynkowej, co ma kluczowe znaczenie w zarządzaniu ryzykiem

³⁶P.J. Brockwell, R.A. Davis, *Time Series: Theory and Methods*, wydanie 2, Nowy Jork 2006, s. 438

portfela i optymalizacji alokacji aktywów. Z powodów problemów natury obliczeniowej, trudno jest uogólnić jeden model MGARCH. Istnieje wiele wersji wielowymiarowego modelu GARCH (np. VEC-GARCH, BEKK-GARCH itd.), jednakże w tej pracy interesować nas będą głównie DCC-GARCH oraz CCC-GARCH. Model DCC-GARCH³⁷ (ang. *Dynamic Conditional Correlation GARCH*) jest dwuetapowym podejściem do modelowania zmienności i korelacji w wielowymiarowym kontekście. Ogólna forma modelu DCC-GARCH może być przedstawiona następująco:

$$\boldsymbol{\varepsilon}_t = \mathbf{H}_t^{1/2} \mathbf{z}_t, \quad (1.50)$$

gdzie:

- $\boldsymbol{\varepsilon}_t$ jest wektorem innowacji (szoków) dla wielu aktywów w chwili t ,
- \mathbf{H}_t jest macierzą wariancji-kowariancji warunkowej w chwili t ,
- $\mathbf{H}_t^{1/2}$ jest pierwiastkiem macierzowym macierzy \mathbf{H}_t , co umożliwia transformację niezależnych innowacji \mathbf{z}_t na innowacje z odpowiednią strukturą korelacji,
- \mathbf{z}_t jest wektorem innowacji, które mają standardowy wielowymiarowy rozkład normalny z średnią równą wektorowi zerowemu i macierzą kowariancji równą macierzy jednostkowej.

Macierz wariancji-kowariancji warunkowej \mathbf{H}_t jest definiowana jako:

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t, \quad (1.51)$$

gdzie:

- \mathbf{D}_t jest macierzą diagonalną zawierającą pierwiastki kwadratowe z wariancji warunkowych estymowanych dla każdego aktywa indywidualnie przy użyciu modelu GARCH,
- \mathbf{R}_t jest macierzą korelacji warunkowej w chwili t , modelowaną w ramach podejścia DCC.

Dynamika macierzy korelacji \mathbf{R}_t w modelu DCC-GARCH jest często opisywana za pomocą specyficznego mechanizmu aktualizacji, który pozwala na modelowanie

³⁷Y. Tse, A. Tsui, *A Multivariate GARCH Model with Time-Varying Correlations*. *Journal of Business and Economic Statistics*, nr. 20, s. 351-362

zmieniających się w czasie korelacji między aktywami. Pierwszy etap polega na indywidualnym modelowaniu wariancji dla każdego aktywa przy użyciu standardowego modelu GARCH. Dla i -tego aktywa w portfelu, model GARCH może być przedstawiony jako:

$$\sigma_{i,t}^2 = \omega_i + \sum_{k=1}^q \alpha_{i,k} \varepsilon_{i,t-k}^2 + \sum_{k=1}^p \beta_{i,k} \sigma_{i,t-k}^2, \quad (1.52)$$

gdzie $\sigma_{i,t}^2$ to wariancja warunkowa aktywa i w czasie t , $\varepsilon_{i,t-k}^2$ to kwadraty innowacji z przeszłości, a ω_i , $\alpha_{i,k}$, i $\beta_{i,k}$ to parametry modelu. Na tej podstawie konstruowana jest macierz diagonalna \mathbf{D}_t , której elementy diagonalne to pierwiastki kwadratowe z estymowanych wariancji warunkowych, czyli $\mathbf{D}_t = \text{diag}(\sqrt{\sigma_{1,t}^2}, \sqrt{\sigma_{2,t}^2}, \dots, \sqrt{\sigma_{n,t}^2})$.

Drugi etap dotyczy modelowania dynamicznej korelacji między aktywami. Mechanizm aktualizacji dla macierzy korelacji \mathbf{R}_t w modelu DCC jest opisany przez równania:

$$\mathbf{Q}_t = (1 - \alpha - \beta) \bar{\mathbf{Q}} + \alpha (\mathbf{u}_{t-1} \mathbf{u}_{t-1}') + \beta \mathbf{Q}_{t-1}, \quad (1.53)$$

$$\mathbf{R}_t = \mathbf{D}_t^{-1} \mathbf{Q}_t \mathbf{D}_t^{-1}, \quad (1.54)$$

gdzie \mathbf{Q}_t to macierz warunkowych kowariancji skalowanych, $\bar{\mathbf{Q}}$ to długookresowa średnia macierzy kowariancji skalowanych, α i β to parametry modelu DCC kontrolujące wpływ poprzednich innowacji i kowariancji na bieżącą macierz \mathbf{Q}_t . Macierz \mathbf{R}_t , będąca znormalizowaną wersją \mathbf{Q}_t , przedstawia warunkowe korelacje między aktywami.

Całościowo model DCC(1,1)-GARCH(p,q), wygląda następująco:

$$\mathbf{z}_t = \mathbf{H}_t^{1/2} \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim IID(0, \mathbf{I}_K), \quad (1.55)$$

gdzie $\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t$ jest macierzą wariancji-kowariancji warunkowej,

$$\mathbf{D}_t = \text{diag}(\sqrt{h_{11,t}}, \dots, \sqrt{h_{KK,t}}), \quad (1.56)$$

$$z h_{ii,t} = \omega_i + \sum_{k=1}^p \alpha_{ii,k} h_{ii,t-k} + \sum_{k=1}^q \beta_{ii,k} z_{i,t-k}^2,$$

$$\mathbf{R}_t = (\mathbf{I}_K - \mathbf{Q}_t)^{1/2} \mathbf{Q}_t (\mathbf{I}_K - \mathbf{Q}_t)^{1/2}, \quad (1.57)$$

$$\mathbf{Q}_t = (1 - \beta - \alpha) \mathbf{S} + \beta \mathbf{u}_{t-1} \mathbf{u}_{t-1}' + \alpha \mathbf{Q}_{t-1}, \quad (1.58)$$

gdzie $\mathbf{u}_t = (u_{1,t}, \dots, u_{K,t})'$ i $u_{i,t} = \frac{z_{i,t}}{\sqrt{h_{ii,t}}}$ dla $i = 1, \dots, K$,

$\beta > 0$, $\alpha > 0$, $0 < \beta + \alpha < 1$,

$\mathbf{S} = [\text{cov}(u_{i,t}, u_{j,t})]_{i,j=1}^K$ jest niezależną od czasu macierzą bezwarunkowej kowariancji standaryzowanych reszt,

\mathbf{R}_t jest macierzą warunkowych korelacji dla wektora \mathbf{u}_t ,

\mathbf{Q}_t jest pomocniczą macierzą wykorzystywaną w definicji macierzy \mathbf{R}_t , będącą macierzą $K \times K$, symetryczną i dodatnio określoną,

\mathbf{S} jest bezwarunkową macierzą kowariancji standaryzowanych reszt. Parametry przy DCC oznaczają odpowiednio: wpływ kwadratów standaryzowanych reszt na aktualną macierz korelacji oraz wpływ poprzedniej macierzy korelacji na aktualną macierz korelacji.

Model CCC-GARCH (ang. *Constant Conditional Correlation GARCH process*) różni się od DCC tym, że zakłada się że korelacje między aktywami są stałe w czasie, przez co wzór ten jest podobny do DCC, jednakże mniej skomplikowany.

W tym podrozdziale opisałem główne dyskretne modele opisu podstawowych cech jakie spełniają aktywa na rynkach finansowych. Cechy te posłużą nam w ostatnim podrozdziale I rozdziału w omawianiu optymalizacji portfelowej. Podobnie jak modele ciągłe są one w stanie opisać zarówno oczekiwane zwroty jak i zmienność, które to są estymatorami zysku jak i ryzyka.

1.3. Sztuczne sieci neuronowe oraz ich zastosowanie w finansach ilościowych

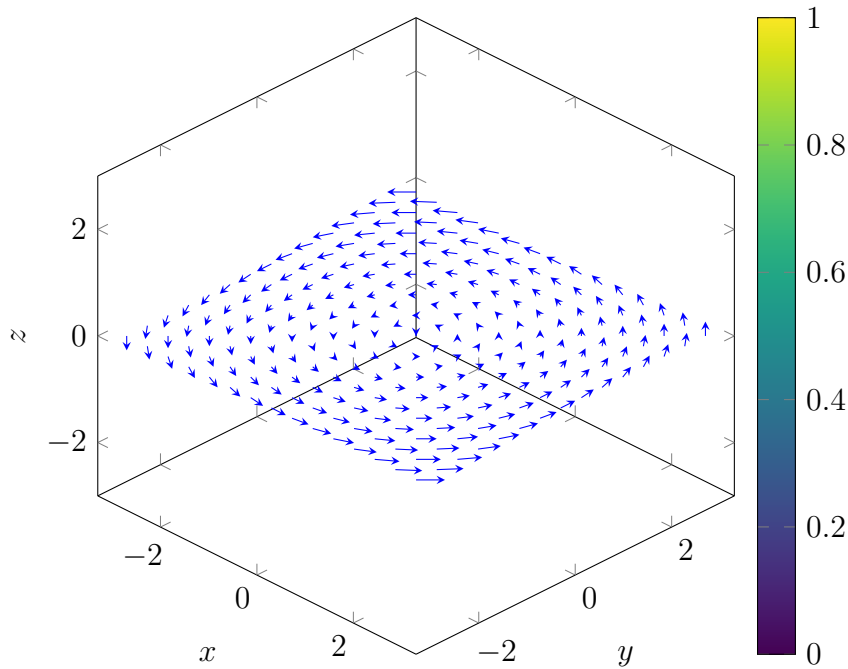
W ostatnich latach jesteśmy świadkami wzrostu znaczenia sztucznej inteligencji w życiu codziennym. Przyczyn tego zjawiska jest wiele, przede wszystkim wiąże się ono z sukcesem implementacji modeli językowych w ramach aplikacji sieciowych takich jak ChatGPT z OpenAI. Która pozwala użytkownikom na całym świecie w różnych językach, otrzymywać odpowiedzi do zapytań stawianych przez użytkowników. Wraz z tym (nie da się ukryć) postępem mnożyły się różne niekiedy absurdalne hipotezy, czym to rzekomo nie jest sztuczna inteligencja. Niemal niekiedy wręcz dochodziło do mylenia samoświadomości z automatem predykcyjnym (czym w istocie jest tzw. sztuczna inteligencja), przez wszelakie indywidua nie wiedzące czym to jest. W konsekwencji internet jest pełen fałszywych autorytetów które pod płaszcem rzekomej nauki, propagowały i nakręcały paranoje społeczną związaną z tzw. sztuczną inteligencją. Żeby zrozumieć na czym tak naprawdę to wszystko polega,

trzeba przebrnąć przez matematykę która się za tym kryje. Najpierw zdefiniujemy pole wektorowe które jest podstawową przestrzenią miar, w tych systemach.

$$\mathbf{F}(x, y, z) = P(x, y, z)\mathbf{i} + Q(x, y, z)\mathbf{j} + R(x, y, z)\mathbf{k} \quad (1.59)$$

gdzie:

- $\mathbf{F}(x, y, z)$ jest polem wektorowym zdefiniowanym w punkcie (x, y, z) przestrzeni trójwymiarowej,
- $P(x, y, z)$, $Q(x, y, z)$, $R(x, y, z)$ są funkcjami skalarnymi przypisującymi wartości w zależności od położenia (x, y, z) w przestrzeni,
- $\mathbf{i}, \mathbf{j}, \mathbf{k}$ są wektorami jednostkowymi wzdłuż osi x, y, z odpowiednio.



Podstawowe operacje na tym polu obejmują różniczkowanie i całkowanie wektorów. Kluczowe z nich to:

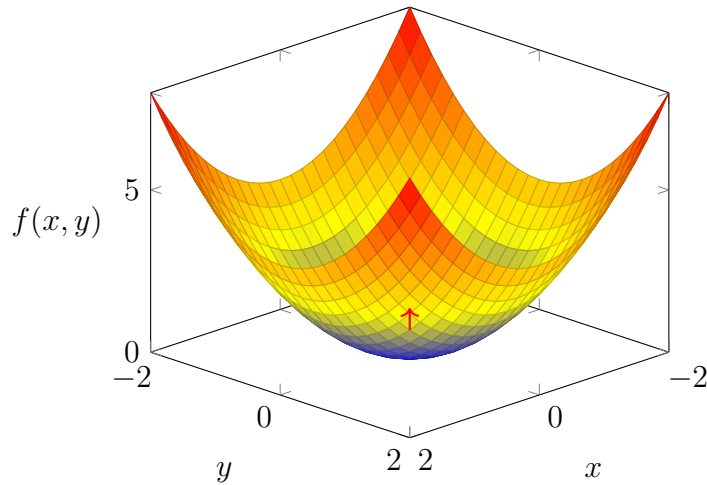
- **Gradient:** jest operacją, która przypisuje każdemu punktowi przestrzeni wektor wskazujący kierunek i szybkość największego wzrostu danej funkcji skalarnej.
- **Dywergencja:** mierzy miarę źródła lub ujścia pola wektorowego w danym punkcie, tj. jak bardzo pole "rozchodzi się" z tego punktu.
- **Rotacja:** określa obrót pola wektorowego wokół danego punktu.

- **Laplasjan:** jest operacją drugiego rzędu, która mierzy różnicę między wartością funkcji w danym punkcie a jej średnią wartością wokół tego punktu.

W kontekście sieci neuronowych, najbardziej interesującą dla nas operacją jest gradient. Gradient funkcji skalarniej $f(x_1, x_2, \dots, x_n)$ w n -wymiarowej przestrzeni euklidesowej jest wektorem wszystkich jej pierwszych cząstkowych pochodnych. Oznacza to, że gradient wskazuje kierunek najszybszego wzrostu funkcji w danym punkcie i jest określony jako:

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right) \quad (1.60)$$

Rozważmy funkcję $f(x, y) = x^2 + y^2$. Jest to prosta funkcja skalarna, której gradient w dowolnym punkcie (x, y) wskazuje kierunek najszybszego wzrostu wartości funkcji.



Wykres 1.2: Wizualizacja gradientu

Gradient jest szczególnie wykorzystywany w uczeniu sieci neuronowych, bowiem szukamy jego minimum dla zadanych wag i biasów w stosunku do funkcji kosztu, w przypadku rysunku 1.2 jest to obszar zacieniowany na ciemnoniebiesko. Kolejnymi szczególnie ważną operacją w sieciach neuronowych a konkretnie w uczeniu głębokim (ang. *Deep Learning*) jest reguła łańcuchowa. Najpierw zdefiniujmy sobie funkcję złożoną $f \circ g$ dwóch funkcji $f : Y \rightarrow Z$ i $g : X \rightarrow Y$ jest definiowana jako:

$$(f \circ g)(x) = f(g(x)) \quad (1.61)$$

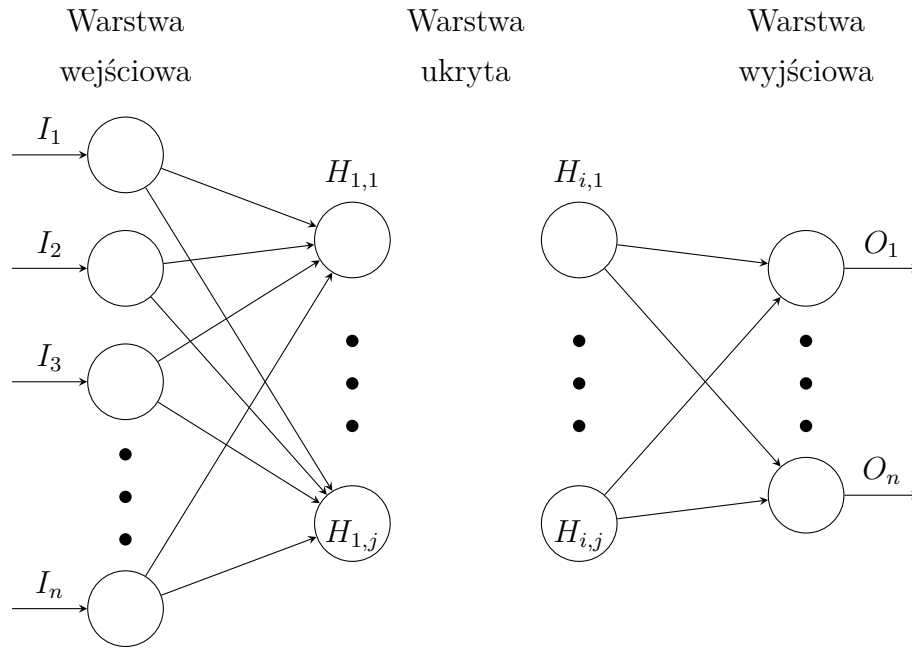
dla każdego $x \in X$. Reguła łańcuchowa służy do obliczania pochodnej funkcji złożonej. Dla funkcji $y = f(u)$ i $u = g(x)$, pochodna funkcji złożonej y względem x jest dana przez:

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx} \quad (1.62)$$

Dla złożenia n funkcji, pochodna funkcji złożonej y względem x obliczana jest jako:

$$\frac{dy}{dx} = \frac{dy}{df_{n-1}} \cdot \frac{df_{n-1}}{df_{n-2}} \cdot \dots \cdot \frac{df_2}{df_1} \cdot \frac{df_1}{dx} \quad (1.63)$$

Znając już wyżej wymienione narzędzia możemy przedstawić schemat działania sieci neuronowych (rysunek poniżej).



Wykres 1.3: Schemat sieci neuronowej z n warstwami ukrytymi, gdzie $H_{i,j}$

Poniżej znajdują się wytłumaczenie zmiennych

- I_n - symbolizuje n -te wejście do sieci neuronowej. Wejścia te są punktem startowym przepływu informacji przez sieć i mogą reprezentować różne cechy danych wejściowych.
- $H_{i,j}$ - oznacza j -ty neuron w i -tej warstwie ukrytej. Neurony te przetwarzają informacje otrzymane z warstwy wejściowej i z poprzednich warstw ukrytych, potencjalnie w bardzo złożony sposób, dzięki nieliniowym funkcjom aktywacji i zdolności do tworzenia złożonych abstrakcji.
- O_n - reprezentuje n -te wyjście sieci, czyli wynik przetwarzania danych wejściowych przez sieć. W zależności od zadania, wyjścia te mogą reprezentować klasy w problemach klasyfikacji lub wartości ciągłe w problemach regresji.

Poniżej znajduje się opis warstw

- **Warstwa wejściowa** jest pierwszą warstwą sieci neuronowej, która odbiera dane wejściowe. Każdy neuron w tej warstwie odpowiada za jedną cechę danych wejściowych.
- **Warstwa ukryta** zawiera neurony, które nie są bezpośrednio widoczne z zewnątrz sieci. Neurony te przetwarzają informacje otrzymane z warstwy wejściowej, tworząc abstrakcyjne reprezentacje danych, które ułatwiają zadanie, np. klasyfikację czy regresję. Głębokość warstwy ukrytej, determinuje liczbę złożań funkcji (przyjmowanie jako argumentów wartości funkcji aktywacyjnej w sieci poprzedniej)
- **Warstwa wyjściowa** dostarcza wynik działania sieci. Liczba neuronów w tej warstwie i rodzaj funkcji aktywacji zależą od specyficznego zadania, które ma realizować sieć.

W najbardziej ogólnej postaci sieć pokazaną na rysunku 1.3 możemy utożsamiać ze złożoną funkcją tensorową. Gdyż tensor jest uogólnieniem tablic wielowymiarowych, macierzy oraz skalarów. Tak więc na samym wstępie przekazywane są dane wejściowe do warstwy ukrytej. Następnie do każdego neuronu w warstwie ukrytej obliczamy wynik funkcji aktywacyjnej, gdzie argumentem tej funkcji jest liniowa kombinacja wag i wejść zsumowane na końcu z biasem.

$$h = f \left(\sum_{i=1}^n w_i x_i + b \right) \quad (1.64)$$

gdzie:

- h jest aktywacją neuronu,
- f jest funkcją aktywacyjną,
- w_i są wagami przypisanymi do każdego z n sygnałów wejściowych x_i ,
- b jest wyrazem wolnym, zwanym również biasem.

Funkcje aktywacyjne decydują o aktywacji neuronów na podstawie przetworzonych sygnałów wejściowych. Poniżej znajduje się lista z najpopularniejszymi funkcjami aktywacyjnymi:

1. **Funkcja sigmoidalna (Sigmoid)** - przekształca wejście na wartości w zakresie od 0 do 1, co jest przydatne, gdy oczekiwane wyjście to prawdopodobieństwo.

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

- 2. ReLU (Rectified Linear Unit)** - obecnie jedna z najczęściej używanych funkcji aktywacyjnych, ze względu na jej skuteczność w uczeniu głębokich sieci.

$$f(x) = \max(0, x)$$

- 3. Tangens hiperboliczny (Tanh)** - przekształca wejście na wartości w zakresie od -1 do 1, co czyni go bardziej użytecznym w niektórych zastosowaniach niż sigmoid.

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

- 4. Softmax** - często używana w ostatniej warstwie sieci neuronowej dla zadań klasyfikacyjnych, przypisując prawdopodobieństwa do różnych klas.

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

gdzie z_i to wartość wejściowa i -tego neuronu w warstwie, a sumowanie odbywa się po wszystkich neuronach w tej warstwie.

Po obliczeniu wartości funkcji z argumentu opisanego w równaniu 1.64, wartości wynikowe przesyłane są dalej w ten sam proces. Oznacza to, że ponownie bierzemy tym razem wyniki tych poprzednich funkcji robimy z nich kombinacje linową nowych wag wraz z dodaniem biasa oraz przekazujemy to jako argument do zadanej funkcji aktywacyjnej dla każdego neuronu w danej warstwie ukrytej. Albo też, przekształcamy zadane nam wartości z tych funkcji jako ostatni argument (również w postaci opisanej równaniem 1.64) do ostatecznej funkcji aktywacyjnej interesującego nas wyniku. Na bazie predykcji teoretycznej porównujemy ją ze stanem rzeczywistym, pomocnym narzędziem tutaj jest funkcja kosztu. W zależności od charakterystyki zmiennej wynikowej, funkcja kosztu może przyjmować różną postać. Najważniejszym aspektem i właściwością jest to że, pozwala ocenić jak dobrze model dopasował się do danych. Dla zadań regresji, gdzie zmienna wynikowa jest ciągła, często stosuje się:

- **Błąd średniokwadratowy (MSE - Mean Squared Error):**

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

gdzie y_i to wartości rzeczywiste, \hat{y}_i to wartości przewidywane przez model, a N to liczba obserwacji.

- **Błąd średni bezwzględny (MAE - Mean Absolute Error):**

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

Dla zadań klasyfikacji, gdzie zmienna wynikowa jest kategoriowa, używa się innych funkcji kosztu, takich jak:

- **Entropia krzyżowa (Cross-Entropy):** dla klasyfikacji binarnej:

$$BCE = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

i dla klasyfikacji wieloklasowej:

$$CE = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log(\hat{y}_{i,c})$$

gdzie $y_{i,c}$ to wskaźnik, czy klasa c jest prawidłową klasą dla obserwacji i , a $\hat{y}_{i,c}$ to przewidywane prawdopodobieństwo, że obserwacja i należy do klasy c , C to liczba klas.

Uzyskując wynik funkcji kosztu, po sprawdzeniu jak model dopasował się do interesującego nas zestawu danych, możemy przejść do najciekawszej części a mianowicie do algorytmu propagacji wstecznej. Algorytm ten służy do obliczenia wrażliwości zmiany każdej wagi w_{ig} oraz biasu b_{ig} w całej sieci neuronowej, na zmianę wartości funkcji kosztu. W istocie algorytm ten jest po prostu realizacją gradientu i reguły łańcuchowej, opisanej równaniem 1.63 oraz 1.60. Bowiem jak wspomniałem wcześniej sieć neuronowa jest tak naprawdę jedynie skomplikowaną funkcją matematyczną, służącą do aproksymacji nieliniowej. Poniżej przedstawiłem kroki w której wykonywana jest pojedyncza realizacja tego algorytmu.

Krok 1: Przeprowadzenie propagacji w przód przez sieć, aby obliczyć wyjście \hat{y} i funkcję kosztu $\mathcal{L}(y, \hat{y})$.

Krok 2: Inicjacja propagacji wstecznej przez obliczenie gradientu funkcji kosztu względem przewidywanego wyjścia sieci:

$$\frac{\partial \mathcal{L}}{\partial \hat{y}}$$

Krok 3: Obliczanie gradientów względem wag i biasów w każdej warstwie, stosując regułę łańcuchową: Dla wagi w i biasu b w warstwie l , gradienty są obliczane jako:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial w} &= \frac{\partial \mathcal{L}}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial a^{(l)}} \cdot \frac{\partial a^{(l)}}{\partial w} \\ \frac{\partial \mathcal{L}}{\partial b} &= \frac{\partial \mathcal{L}}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial a^{(l)}} \cdot \frac{\partial a^{(l)}}{\partial b} \end{aligned}$$

gdzie $a^{(l)}$ to aktywacja w warstwie l przed zastosowaniem funkcji aktywacji, a $\frac{\partial \hat{y}}{\partial a^{(l)}}$ to pochodna wyjścia sieci względem aktywacji tej warstwy.

Krok 4: Aktualizacja parametrów sieci neuronowej:

$$w := w - \eta \frac{\partial \mathcal{L}}{\partial w}$$

$$b := b - \eta \frac{\partial \mathcal{L}}{\partial b}$$

gdzie η jest współczynnikiem uczenia w praktyce przyjmuje wartości między 0,0001 a 0,1.

Cały proces jest powtarzany do momentu aż interesujące nas mierniki (np. F1, MSE itd.) będą miały zadowalającą dla nas precyzję. Zazwyczaj startujemy z losowymi wartościami parametrów, następnie aktualizujemy je w oparciu o wynik gradientu i wartość współczynnika uczenia, a następnie iterujemy ten proces określoną ilość razy. Wzór który podałem na aktualizację parametrów jest prawdziwy, ale tylko dla gradientu prostego (dla całego zestawu danych). W praktyce ze względu na dosyć duży koszt obliczeniowy nie zawsze się stosuje akurat tą metodę. Poniżej znajduje się lista najpopularniejszych algorytmów gradientowych.

1. Gradient Prostý (Batch Gradient Descent)

Aktualizacja parametrów:

$$\theta = \theta - \eta \cdot \nabla_{\theta} J(\theta)$$

2. Stochastyczny Gradient Prostý (SGD)

Aktualizacja parametrów dla każdej próbki:

$$\theta = \theta - \eta \cdot \nabla_{\theta} J(\theta; x^{(i)}, y^{(i)})$$

3. Mini-Batch Gradient Descent

Aktualizacja parametrów dla każdego mini-batcha:

$$\theta = \theta - \eta \cdot \nabla_{\theta} J(\theta; X_{\text{mini-batch}}, Y_{\text{mini-batch}})$$

4. Momentum

Aktualizacja parametrów z wykorzystaniem momentu:

$$v_t = \gamma v_{t-1} + \eta \cdot \nabla_{\theta} J(\theta)$$

$$\theta = \theta - v_t$$

5. Nesterov Accelerated Gradient (NAG)

Aktualizacja parametrów z wyprzedzeniem:

$$v_t = \gamma v_{t-1} + \eta \cdot \nabla_{\theta} J(\theta - \gamma v_{t-1})$$

$$\theta = \theta - v_t$$

6. Adagrad

Aktualizacja parametrów z adaptacyjnym współczynnikiem uczenia:

$$\theta = \theta - \frac{\eta}{\sqrt{G_t + \epsilon}} \odot \nabla_{\theta} J(\theta)$$

gdzie G_t to diagonalna macierz sumująca kwadraty gradientów do czasu t , a ϵ to mała stała zapobiegająca dzieleniu przez zero.

7. RMSprop

Aktualizacja parametrów z wygładzonymi kwadratami gradientów:

$$\theta = \theta - \frac{\eta}{\sqrt{E[g^2]_t + \epsilon}} \odot \nabla_{\theta} J(\theta)$$

8. Adam (Adaptive Moment Estimation)

Aktualizacja parametrów z momentami pierwszego i drugiego rzędu:

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) \nabla_{\theta} J(\theta) \\ v_t &= \beta_2 v_{t-1} + (1 - \beta_2) (\nabla_{\theta} J(\theta))^2 \\ \hat{m}_t &= \frac{m_t}{1 - \beta_1^t}, \quad \hat{v}_t = \frac{v_t}{1 - \beta_2^t} \\ \theta &= \theta - \frac{\eta}{\sqrt{\hat{v}_t + \epsilon}} \odot \hat{m}_t \end{aligned}$$

Zasadniczo co może się wydawać zaskoczeniem, jedną z pierwszych aplikacji sieci neuronowych było właśnie modelowanie rynków kapitałowych, już w latach 90 około 3 lata po odkryciu algorytmu propagacji wstecznej przez Yana LeCuna. A więc te metody wcale nie są innowatorskie na tej płaszczyźnie. Jest to kompletnie odmienne podejście, od tych które prezentowałem w poprzednich podrozdziałach, jednakże posiadające również silny potencjał prognostyczny. W kontekście modelowania zmienności cen aktywów, interesować nas będzie specyficzny rodzaj sieci neuronowych dedykowany dla szeregów czasowych. Konkretnie chodzi tu o Rekurencyjne Sieci Neuronowe (RNN), które poza pewnymi szczegółami w ogólności działają tak samo jak klasyczna sieć neuronowa. W tym sensie, że parametry również podlegają propagacji wstecznej, a argumenty są funkcją wyników poprzednich warstw. Teraz omówmy pojęcie rekurencyjnych sieci neuronowych

1.4. Sterowanie stochastyczne

1.5. Arbitraż statystyczny

1.6. Teoria optymalizacji portfela

,

Rozdział 2

Empiryczna budowa portfela

2.1. Założenia i strategia inwestycyjna

2.2. Podejście stochastyczne

2.2.1. Dobór aktywów i wag portfela

2.2.2. Modelowanie równań momentów

2.2.3. Analiza predykcyjna

2.3. Podejście algorytmiczne

2.3.1. Metody heurystyczne w doborze portfela

2.3.2. Algorytmy ewolucyjne w optymalizacji wag aktywów

2.3.3. Zastosowanie sieci neuronowych do identyfikacji wzorców rynkowych

2.3.4. Zastosowanie analizy skupień do segmentacji rynku

2.3.5. Algorytmy do oceny i zarządzania ryzykiem portfela

2.3.6. Automatyzacja procesów inwestycyjnych za pomocą algorytmów

2.3.7. Integracja zewnętrznych danych do analizy algorytmicznej

Rozdział 3

Wyniki oraz podsumowanie

3.1. Wynik finansowy portfeli

3.2. Odporność systemów w czasie

3.3. Podsumowanie

Spis tabel

Spis rysunków

1.1	Proces Wienera-Bacheliera, opracowanie własne	9
1.2	Wizualizacja gradientu	30
1.3	Schemat sieci neuronowej z n warstwami ukrytymi, gdzie $H_{i,j}$	31

Spis kodów źródłowych