Contents

1 formula

First we only consider **one play**

$$G(P_n, w^1) = \sum_{i=1}^{S} \tilde{\theta}_i \tanh(\theta_i P_n + \theta_{i0}) + \tilde{\theta_0} \tag{1}$$

Where $P_n = [p_1, p_2, p_3, ..., p_n]$, $G(P_n, w^1) = [y_1, y_2, y_3, ..., y_n]$
$\forall i \in [1, ..., S]$, $\theta_i P_n = (\theta_i p_1, \theta_i p_2, \theta_i p_3, ..., \theta_i p_n)$,
So $\tanh(\theta_i P_n + \theta_{i0}) = [\tanh(\theta_i p_1 + \theta_{i0}), \tanh(\theta_i p_2 + \theta_{i0}), \tanh(\theta_i p_3 + \theta_{i0}), ..., \tanh(\theta_i p_n + \theta_{i0})]$
So $\sum_{i=1}^{S} \tilde{\theta}_i \tanh(\theta_i P_n + \theta_{i0}) + \tilde{\theta_0} = [\sum_{i=1}^{S} \tilde{\theta}_i \tanh(\theta_i p_1 + \theta_{i0}) + \tilde{\theta_0}, \sum_{i=1}^{S} \tilde{\theta}_i \tanh(\theta_i p_2 + \theta_{i0}) + \tilde{\theta_0}, ..., \sum_{i=1}^{S} \tilde{\theta}_i \tanh(\theta_i p_n + \theta_{i0}) + \tilde{\theta_0}] = [y_1, y_2, ..., y_n]$.
Take y_j, where $j \in [1, ..., n]$ for example
Let $z_j = \theta_i p_j + \theta_{i0}$ and $f(z_j) = \tanh(\theta_i p_j + \theta_{i0})$, we obtain

$$\begin{aligned} y_j &= \sum_{i=1}^{S} \tilde{\theta}_i \tanh(\theta_i p_j + \theta_{i0}) + \tilde{\theta_0} && (2) \\ &= \sum_{i=1}^{S} \tilde{\theta}_i f(z_j) + \tilde{\theta_{i0}} && (3) \\ & && (4) \end{aligned}$$

Calculate derivation for y_j,

$$\begin{aligned} \frac{\partial y_j}{\partial p_j} &= \sum_{i=1}^{S} \tilde{\theta}_i \theta_i \frac{\partial f(z_j)}{\partial z_j} && (5) \\ & && (6) \end{aligned}$$

Now let's consider the mapping between p_j and x_j. let $\sigma_j = w^1 x_j - p_{j-1}$

$$p_j = \Phi(\sigma_j) + p_{j-1} \tag{7}$$

and

$$\Phi(x) = \begin{cases} x, x > 0 \\ 0, -1 < x < 0 \\ x - 1, x < -1 \end{cases}$$

(8)

Using chain rule, we obtain

$$\begin{aligned} \frac{\partial y_j}{\partial x_j} &= \frac{\partial y_j}{\partial p_j} \frac{\partial p_j}{\partial x_j} & (9) \\ &= \sum_{i=1}^{S} \tilde{\theta}_i \theta_i w^1 \frac{\partial f(z_j)}{\partial z_j} \frac{\partial \Phi(\sigma_j)}{\partial \sigma_j} & (10) \end{aligned}$$

To consider **multiple plays** case, we reformulate the derivation as following:

$$\begin{aligned} \frac{\partial y_j^1}{\partial x_j} &= \frac{\partial y_j^1}{\partial p_j^1} \frac{\partial {p_j}^1}{\partial x_j} & (11) \\ &= \sum_{i=1}^{S} \tilde{\theta_i^1} \theta_i^1 w^1 \frac{\partial f(z_j^1)}{\partial z_j^1} \frac{\partial \Phi(\sigma_j^1)}{\partial \sigma_j^1} & (12) \end{aligned}$$

Now from the architecture, we know that if we have P plays,

$$F = \frac{1}{P} \sum_{k=1}^{P} G^k \tag{13}$$

Where $F = [f_1, f_2, ..., f_n]$, and

$$f_j = \frac{1}{P} \sum_{k=1}^{P} y_j^k \tag{14}$$

our derivation is:

$$\frac{\partial f_j}{\partial x_j} = \frac{1}{P}\sum_{k=1}^{P}\frac{\partial y_j^k}{\partial x_j} \tag{15}$$

$$= \frac{1}{P}\sum_{k=1}^{P}\frac{\partial y_j^k}{\partial p_j^k}\frac{\partial p_j^k}{\partial x_j} \tag{16}$$

$$= \frac{1}{P}\sum_{k=1}^{P}\sum_{i=1}^{S}\tilde{\theta}_i^k\theta_i^k w^k\frac{\partial f(z_j^k)}{\partial z_j^k}\frac{\partial \Phi(\sigma_j^k)}{\partial \sigma_j^k} \tag{17}$$

(18)

2 RNN gradients

$$\frac{\partial p_j}{\partial x_j} = \Phi'(\sigma_j)\frac{\partial \sigma_j}{\partial x_j} \tag{19}$$

$$= \Phi'(\sigma_j)w^1 \tag{20}$$

(21)

(22)

$\partial p_{j+1}\overline{\partial x_j = \frac{\partial(\Phi(\sigma_{j+1})+p_j)}{\partial x_j}}$(23)

$= \Phi'(\sigma_{j+1})\frac{\partial \sigma_{j+1}}{\partial x_j} + \frac{\partial p_j}{\partial x_j}$(24)

$= \Phi'(\sigma_{j+1})\frac{\partial(w^1 x_{j+1} - p_j)}{\partial x_j} + \frac{\partial p_j}{\partial x_j}$(25)

$= (1\text{-}\Phi'(\sigma_{j+1}))\Phi'(\sigma_j)w^1$

(26)

(27)

$\partial p_{j+2}\overline{\partial x_j = (1-\Phi'(\sigma_{j+2}))(1-\Phi'(\sigma_{j+1}))\Phi'(\sigma_j)w^1}$(28)

(29)

(30)

$\partial p_{j+i}\overline{\partial x_j = (1-\Phi'(\sigma_{j+i}))\ldots(1-\Phi'(\sigma_{j+1}))\Phi'(\sigma_j)w^1}$(31)

(32)

3 implementation

Consider $P_n = [p_1, p_2, p_3, ..., p_n]$, $G(P_n, w^1) = [y_1, y_2, y_3, ..., y_n]$

4 Trading

Assume, we observe prices $p_1, p_2, ..., p_N$ for a fixed $N > 0$. Based on Dima's paper, assume that the price p_n hysteretically depends on the underlying noise b_n, with b_n being a Brownian motion. Using the notation

$$\mathcal{B}_n := (b_1, b_2, ..., b_n), \mathcal{P}_n := (p_1, p_2, ..., p_n)$$

we have

$$b_0 = 0,\, b_n \sim \mathcal{N}(b_{n-1} + \mu_b, \sigma_b) \tag{33}$$

$$p_n = F(\mathcal{B}_n, W_p) \tag{34}$$

Based on Dima's paperr again, the underlying noise b_n can be expressed as a hysteresis operator depending on the observed prices p_n, i.e.,

$$b_n = G(\mathcal{P}_n, W_b)$$

This is very nice as long as F and G are Prandtl-Ishlinskii operators. However, if one explicitly adds N's strategy, G becomes Preisach and F is not Preisach anymore. It is not clear how well it can be approximated by compositions of plays and nonlinear functions

5 Direct Learning

We learn the parameters W_b, μ_b, σ_b and the initial state p_0 of the network G by maximizing the likelihood of $\mathcal{P}$. Since $\mathcal{P}$ is the determenistic function of a random variable $mathcal B$, its probability density is given by

$$\begin{aligned} p(\mathcal{P}) &= p(p_1, p_2, ..., p_N) && (35)\\ &= p_b(b_1, b_2, ..., b_N)\, |\det \mathcal{J}(\mathcal{P})| && (36)\\ &= p_b(G(\mathcal{P}_1, W_b), G(\mathcal{P}_2, W_b), ..., G(\mathcal{P}_N, W_b))\, |\det \mathcal{J}(\mathcal{P})| && (37)\\ & && (38) \end{aligned}$$

where

$$\begin{aligned}
p_b(\mathcal{B}) &= \prod_{n=1}^{N} p_b(b_1, b_2, ..., b_N) && (39)\\
&= \prod_{n=1}^{N} p_b(b_n|b_{n-1}) && (40)\\
&= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(b_n - b_{n-1} - \mu_b)^2}{2\sigma^2}\right) && (41)
\end{aligned}$$

is the probability distribution of $\mathcal{B}$ and $\mathcal{J}(\mathcal{P})$ is the Jacobian matrix. Recall that G has the causality property, hence $\mathcal{J}(\mathcal{P})$ is a triangular matrix. Therefore, yield

$$\begin{aligned}
p(\mathcal{P}) &= p_b(G(\mathcal{P}_1, W_b), G(\mathcal{P}_2, W_b), ..., G(\mathcal{P}_N, W_b)) \left|\det \mathcal{J}(\mathcal{P})\right| && (42)\\
&= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(b_n - b_{n-1} - \mu_b)^2}{2\sigma^2}\right) \prod_{n=1}^{N} \left|\frac{\partial b_n}{\partial p_n}\right| && (43)\\
&= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(b_n - b_{n-1} - \mu_b)^2}{2\sigma^2}\right) \left|\frac{\partial b_n}{\partial p_n}\right| && (44)
\end{aligned}$$

Thus, maximizing the log-likelihood of $p(\mathcal{P})$ is equivalent to the following:

$$\begin{aligned}
L &\sim \sum_{n=1}^{N} \left(-\frac{(b_n - b_{n-1} - \mu_b)^2}{2\sigma_b^2} + \ln\left|\frac{\partial b_n}{\partial p_n}\right|\right) && (45)\\
&= -\frac{1}{2}\sum_{n=1}^{N} \left[\left(\frac{b_n - b_{n-1} - \mu_b}{\sigma_b}\right)^2 - 2\ln\left|\frac{\partial b_n}{\partial p_n}\right|\right] && (46)
\end{aligned}$$

It's also equivalent to mimizing the loss function as following:

$$\min L = \min \sum_{n=1}^{N} \left[\left(\frac{b_n - b_{n-1} - \mu_b}{\sigma_b}\right)^2 - 2\ln\left|\frac{\partial b_n}{\partial p_n}\right|\right] \quad (47)$$