Mathematics for New Technologies in Finance

Solution sheet 5

Exercise 5.1 (Self financing portfolio) Recall the definition of self financing trading strategy ξ and its associated discounted value process $V = (V_t)_{t=0,\dots,T}$ is given by

$$V_0 := \xi_1 \cdot X_0 \quad \text{and} \quad V_t := \xi_t \cdot X_t \quad \text{for } t = 1, \dots, T.$$

The gains process associated with ξ is defined as

$$G_0 := 0 \quad \text{and} \quad G_t := \sum^{t} \xi_k \cdot (X_k - X_{k-1}) \quad \text{for } t = 1, \dots, T$$

(a) Prove $\xi_t \cdot X_t = \xi_{t+1} \cdot X_t$ for $t = 1, ..., T-1$.

(b) Prove $V_t = V_0 + G_t = \xi_1 \cdot X_0 + \sum_{k=1}^{t} \xi_k \cdot (X_k - X_{k-1})$ for all t.

Solution 5.1

(a) By definition we have

$$\xi_t \cdot S_t = \xi_{t+1} \cdot S_t \text{ for } t = 1, ..., T-1,$$

S_t is the price of the asset at time t. By dividing both sides by S_t^0, we can prove $\xi_t \cdot X_t = \xi_{t+1} \cdot X_t$ for $t = 1, ..., T-1$.

(b) Since (a) holds, we have

$$\xi_{t+1} \cdot X_{t+1} - \xi_t \cdot X_t = \xi_{t+1} \cdot X_{t+1} - \xi_{t+1} \cdot X_t = \xi_{t+1} \cdot (X_{t+1} - X_t)$$

for $t = 1, ..., T-1$, and it's identical to (b).

Exercise 5.2 (Backpropogation of neural network) Let $\theta = (w, b, a) \in \mathbb{R}^3$ and let σ be the activation function. We consider the shallow neural network $f_\theta\colon \mathbb{R} \to \mathbb{R}$ s.t.

$$f_\theta(x) = a\sigma(wx + b).$$

Then we solve the regression problem with 3 data point $(x_i, y_i) \in \mathbb{R}^2$, $i = 1, 2, 3$ by minimizing the L^2 loss

$$\mathcal{L}_f = \sum_{i=1,2,3} (f_\theta(x_i) - y_i)^2.$$

(a) When solving the regression, do we compute $\nabla_{x_0} \mathcal{L}_f$ or $\nabla_\theta \mathcal{L}_f$?

(b) Compute $\partial_w f$ and $\partial_b f$ by chain rule. Do you find any intermediate value computed twice in both $\partial_w f$ and $\partial_b f$?

(c) Consider regression problem as a constrained optimization problem

$$\begin{aligned} &\min \sum_{i=1,2,3} l_i \\ &l_i = (\tilde{y}_i - y_i)^2 \\ &\tilde{y}_i = a\sigma(z_i), \qquad i = 1, 2, 3. \\ &z_i = wx_i + b \end{aligned}$$

Solve it by Lagrange multiplier and relate this with backpropagation.

(d) Generalize this idea to deep neural networks.

Solution 5.2

(a) $\nabla_\theta \mathcal{L}_f$.

(b) Let $z = wx_0 + b$ then

$$\partial_w \mathcal{L}_f = \partial_z \mathcal{L}_f \cdot x_0 = (a\sigma(w_0 x + b) - y_0)\sigma' a(wx_0 + b)x_0,$$
$$\partial_b \mathcal{L}_f = \partial_z \mathcal{L}_f \cdot 1 = (a\sigma(wx_0 + b) - y_0)a\sigma'(wx_0 + b)$$

(c) Consider the Lagrangian

$$\mathcal{L} = l - \lambda_l(l - (y - y_0)^2) - \lambda_y(y - a\sigma(z)) - \lambda_z(z - (wx_0 + b))$$

Compute the gradient

$$\partial_l \mathcal{L} = 1 - \lambda_l$$
$$\partial_y \mathcal{L} = \lambda_l \frac{\partial (y - y_0)^2}{\partial y} - \lambda_y$$
$$\partial_z \mathcal{L} = \lambda_y \frac{\partial a\sigma(z)}{\partial z} - \lambda_z$$
$$\partial_w \mathcal{L} = \lambda_z \frac{\partial (wx_0 + b)}{\partial w}$$
$$\partial_b \mathcal{L} = \lambda_z \frac{\partial (wx_0 + b)}{\partial b}$$

Lettin $\nabla \mathcal{L} = 0$, we get exactly the backpropagation formula.

(d) See [1].

Exercise 5.3 (Backpropagation and cODEs) Translate a one layer neural network to a controlled ODE:

$$L^{(i)} : x \mapsto W^{(i)}x + a^{(0)} \mapsto \phi(W^{(1)}x + a^{(0)}),$$

with a cadlag control $u(t) = 1_{[1,2)}(t) + 2_{[2,\infty)}(t)$ and a time-dependent vector field

$$V(t,x) = 1_{[0,1)}(t)\left(L^{(0)}(x) - x\right) + 1_{[1,\infty)}(t)\left(L^{(1)}(x) - x\right).$$

The corresponding neural network at 'time' 3 is

$$x \mapsto L^{(0)}(x) \mapsto \phi(W^{(1)}L^{(0)}(x) + a^{(1)}).$$

(a) What is the evolution operator $J_{s,3}$?

(b) Calculate the derivative of the network with respect to parameters $W^{(1)}$ and $a^{(1)}$.

Solution 5.3

(a) $J_{s,3}v = v + 1_{[0,1[}(s)\left(dL^{(1)}(X_{s-})\, dL^{(0)}(x)v - v\right) + 1_{[1,2[}(s)\left(dL^{(1)}(X_{s-})\, v - v\right)$

(b)

$$\frac{\partial V}{\partial W^{(1)}} = 1_{[0,1)}(t)\frac{\partial L^{(0)}}{\partial W^{(1)}} + 1_{[1,\infty)}(t)\frac{\partial L^{(1)}}{\partial W^{(1)}} = 1_{[1,\infty)}(t)\frac{\partial L^{(1)}}{\partial W^{(1)}} = 1_{[1,\infty)}(t)W^{(1)}\frac{\partial}{\partial W^{(1)}}\phi(W^{(1)}x + a^{(1)}).$$

Similarly,

$$\frac{\partial V}{\partial a^{(1)}} = 1_{[1,\infty)}(t) W^{(1)} \frac{\partial}{\partial a^{(1)}} \phi(W^{(1)}x + a^{(1)}).$$

Since we have

$$\partial X_T^\theta = \sum_{i=1}^d \int_0^T J_{s+,T} \partial V_i^\theta \left(s-, X_{s-}^\theta\right) du^i(s)$$

We have

$$\partial X_3^{W^{(1)}} = \sum_{i=1}^d \int_1^3 J_{s+,3} \frac{\partial V}{\partial W^{(1)}} du^i(s) \text{ and } \partial X_3^{a^{(1)}} = \sum_{i=1}^d \int_1^3 J_{s+,3} \frac{\partial V}{\partial a^{(1)}} du^i(s)$$

Exercise 5.4 (Hedging) See notebook 1.

Solution 5.4 See solution notebook 1.

Exercise 5.5 (Path-depedent derivatives in a BS market) See notebook 2.

Solution 5.5 See solution notebook 2.

References

[1] Yann LeCun, D Touresky, G Hinton, and T Sejnowski. A theoretical framework for back-propagation. 1:21–28, 1988.