Предмет: RL
Исполнительница: Смалкина Ю.А
Теор. часть Д3 №1.

№1.

Сумма TD(λ) обновляется

$G_t^\lambda = (1-\lambda)\sum_{n=1}^{\infty} \lambda^{n-1} G_{t:t-1} = R_{t+1} + \gamma G_{t+1}^\lambda + \dots$

$G_t^\lambda - \varphi = \dots = \sum \xi \delta_k$

пусть $\gamma^{m+t} \lambda^{k-t}_{m=t} = \xi$ и $\hat{v}(S_t, w) = \varphi$

$$\sum_{t=0}^{\infty} \alpha [(R_t + \gamma \hat{v}(S_{t+1}, w) - \hat{v}(S_t, w)]_{z_t} = \sum_{t=0}^{\infty} \alpha \delta_t [\gamma \lambda z_{t-1} + \nabla\varphi]$$

$$= \sum \alpha \delta_t [\gamma \lambda (\gamma \lambda z_{t-2} + \nabla\varphi + \nabla \hat{v}(S_t, w))] = \dots = \sum_{t=0}^{\infty} \alpha \delta_t [\sum_{m=0}^{t} \gamma^{t-m} \nabla\varphi]$$

т.ч t: $\forall \nabla\varphi$ встреч только одно в $\sum_{m=0} \gamma^{t-m} \lambda^{t-m}$ для $t \geq m$

Теперь, для всех разн t в уникал. S_t:

$$\sum_{t=0}^{\infty} \alpha \delta_t [\sum \gamma^{k-m} \lambda^{m-t}] \nabla\varphi$$

аналогично для δ_t где $m \geq t$

$$\sum_{m=0}^{\infty} \alpha [\sum \xi \delta_m] \nabla\varphi = \sum \alpha [G_t^\lambda - \varphi] \nabla\varphi$$

№2.

① Т1: (эквивалентность между forward и backward views)

Расcм любой forward view с обновлением цели y_k^t:

$\theta_{k+1}^t = \theta_k^t + \eta_k (y_k^t - \phi_k^T \theta_k^t) \phi_k + x_k, \quad 0 \leq k \leq t$

где $\theta_0^t = \theta_0$ для какого-то θ_0 и где $x_k \in \mathbb{R}^n$ - любой в-р, не завис от t.

Теперь предпол, что $y_k^{t+1} - y_k^t$ можно представ

$y_k^{t+1} - y_k^t = c_k (y_{k+1}^{t+1} - y_{k+1}^t) \quad \forall k < t$ (*), c_k - скаляр, не завис от t

Тогда, в итоге веса θ_t^t $\forall t$ ~ θ_t по опр-но $e_0 = \eta_0 \phi_0$ и backward view

$\theta_{t+1} = \theta_t + (y_t^{t+1} - y_t^t) e_t + \eta_t (y_t^t - \varphi_t^T \theta_t)\varphi_t + x_t$,

$e_t = c_{t-1} e_{t-1} + \eta_t (1 - c_{t-1}\varphi_t^T e_{t-1})\varphi_t$, $t > 0$

Д-во: матрица $F_t \doteq I - \eta_t \varphi_t \varphi_t^T$, т.к $\theta^t_{k+1} = F_k \theta^t_k + \eta_k y^t_k \varphi_k$

Вычтем θ_t^t от θ^{t+1}_{t+1} чтобы найти изменение t.

$\theta^{t+1}_{t+1} - \theta^t_t = F_t\theta^{t+1}_t - \theta^t_t + \eta_t y_t^{t+1}\varphi_t + x_t = \dots = F_t(\theta^{t+1}_t - \theta^t_t) + \eta_t(y_t^{t+1} - \theta_t^t \varphi_t^T)\varphi_t + x_t$

теперь рассмотрим $\theta^{t+1}_t - \theta^t_t = F_{t-1}(\theta^{t+1}_{t-1} - \theta^t_{t-1}) + \eta_{t-1}(y^{t+1}_{t-1} - y^t_{t-1})\varphi_{t-1} = F_{t-1}F_{t-2}(\theta^{t+1}_{t-1} - \theta^t_{t-1}) + \eta_{t-2}(y^{t+1}_{t-2} - y^t_{t-2}) F_{t-1}\varphi_{t-2} + \eta_{t-1}(y^{t+1}_{t-1} - y^t_{t-1})\varphi_{t-1} = \dots$ пока не дойдем до $\theta_0^{t+1} - \theta_0^t = 0$ =

$= F_{t-1}\cdots F_0(\theta_0^{t+1} - \theta_0^t) + \sum_{k=0}^{t-1}\eta_k F_{t-1}\cdots F_{k+1}(y_k^{t+1} - y_k^t)\varphi_k$

$= \dots = \sum_{k=0}^{t-1} \eta_k F_{t-1}\cdots F_{k+1} c_k (y^{t+1}_{k+1} - y^t_{k+1})\varphi_k$ (из (*))

$= \underbrace{c_{t-1} \sum_{k=0}^{t-1}\eta_k \left(\prod_{j=k}^{t-2} c_j\right) F_{t-1}\cdots F_{k+1}\varphi_k}_{\doteq e_{t-1}} (y_t^{t+1} - y_t^t) = c_{t-1}e_{t-1}(y_t^{t+1} - y_t^t)$

и e_t можно вычисл рекур.

$\theta^{t+1}_{t+1} - \theta^t_t = \dots = (y_t^{t+1} - y_t^t)e_t + \eta_t(y_t^t - \varphi_t^T\theta_t)\varphi_t + x_t$.

т.к получилось $\theta_{0,t} \doteq \theta_0\ \forall t$ по индукции все д-ся. □

③ w_t^t опр forward view ~ w_t опр backward view

$e_t^w = \rho_{t-1}\gamma_t\lambda_t e^w_{t-1} + \beta_t(1 - \rho_{t-1}\gamma_t\lambda_t \varphi_t^T e^w_{t-1})\varphi_t$,

$w_{t+1} = w_t + \rho_t \delta_t e_t^w - \beta_t \varphi_t^T w_t \varphi_t$ где $e_0^w = \beta_0\varphi_0$, $w_0 = w_0^t\ \forall t$.

и $\delta_t \doteq R_{t+1} + \gamma \varphi_{t+1}^T\theta_t - \theta_t\varphi_t^T$

Д-во представим $\theta_t = w_t$, $\eta_t = \beta_t$, $x_t = 0$ и $y^t_k = \delta^{\lambda\rho}_{k,t}$

тогда

$$\delta^{\lambda\rho}_{k,t+1} - \delta^{\lambda\rho}_{k,t} = \rho_k \gamma_{k+1} \lambda_{k+1} (\delta^{\lambda\rho}_{k+1,t+1} - \delta^{\lambda\rho}_{k+1,t})$$

$\Rightarrow C_k = \rho_k \gamma_{k+1} \lambda_{k+1}$, $y^t_t = \delta^{\lambda\rho}_{t,t} = 0$ и $y^{t+1}_t - y^t_t = \rho_t \delta_t$ $\square$

3 теперь рассм, что $\theta^t_t \sim \theta_t$.

$$e_t = \rho_t(\gamma_t \lambda_t e_{t-1} + \alpha_t(1 - \rho_t \gamma_t \lambda_t \phi_t^T e_{t-1})\phi_t)$$

$$e^\nabla_t = \rho_t(\gamma_t \lambda_t e_{t-1} + \phi_t)$$

$$\theta_{t+1} = \theta_t + \delta_t e_t + (e_t - \alpha_t \rho_t \phi_t)(\theta_t - \theta_{t-1})^T \phi_t - \alpha_t \gamma_{t+1}(1 - \lambda_{t+1}) \omega_t^T e^\nabla_t \phi_{t+1}$$

Полное д-во я приводить не буду, обозначу только основные моменты опр-им η_t, x_k, y^t_k и последнее y^{t+1}_k, разницу y^{t+1}_k и y^t_t применяя Т1. распишем θ_{t+1} через $\theta_t + \ldots (e_t, \alpha_t, \phi_t, \rho_t, \omega_t)$

Тогда можно опр алгоритм ~~true~~ online GTD(λ)

Рассм голландские (dutch) следы e_t и e^ω_t, e^∇_t следует из коррек градиента. И они явл взаимозамен.

Т.е точн GTD(0): $\lambda = 0$:

$$\theta_{t+1} = \theta_t + \alpha_t \rho_t \delta_t \phi_t - \alpha_t \rho_t \gamma_{t+1} \omega_t^T \phi_t \phi_{t+1}$$

$$\omega_{t+1} = \omega_t + \beta_t \rho_t \delta_t \phi_t - \beta_t \phi_t^T \omega_t \phi_t$$