Stohastična optimizacija v diskretnem času

avtor: Darjan Pavšič
mentor: izred. prof. dr. Mihael Perman

Kratka predstavitev
4. december 2019

Vsebina

- Optimalna kontrola in deterministično dinamično programiranje
- Zgled determinističnega dinamičnega programiranja
- Ideja vpeljave na stohastično dinamično programiranje

Optimalna kontrola in dinamično programiranje

Predpostavke za problem optimalne kontrole:

- Imamo diskreten čas, torej $t \in \mathbb{N}_0$.
- Ekonomija je skozi čas opisana z dvema spremenljivkama: stanje x_t in kontrolna spremenljivka u_t
- Vemo, kakšno je začetno stanje x_0, prav tako je podan razvoj stanja preko funkcije kontrolne spremenljivke: $x_{t+1} = g_t(x_t, u_t)$
- Imamo množico m vrednosti kontrolne spremenljivke za vsako obdobje $t \in \mathbb{N}_0$, $u_t \in \mathcal{U} \subset \mathbb{R}^m$

Optimalna kontrola in dinamično programiranje

- Torej se za vsako zaporedje kontrol $u \equiv \{u_0, u_1, \ldots, u_t\} \in \mathcal{U}$ ekonomija lahko giba po velikem številu dopustnih poti $x \equiv \{x_0, x_1, \ldots, \}$, ki so določene z $x_{t+1} = g_t(x_t, u_t), u_t \in \mathcal{U}$

- Če imamo podan kriterij, na podlagi katerega lahko ovrednotimo vse dopustne poti z vrednostjo $U(x_0, x_1, \ldots, u_0, u_1, \ldots)$ in obstaja optimalna kontrola $u^* = \{u_0^*, u_1^*, ...\}$, ki maksimizira U, potem obstaja tudi optimalna pot za spremenljivko stanja $x^* \equiv \{x_0, x_1^*, \ldots\}$.

Optimalna kontrola in dinamično programiranje

Predvsem v zvezi z ekonomskimi in finančnimi problemi imamo še dodatne predpostavke:

- x_t je lahko npr. vrednost delnice, ki se meri ob začetku obdobja, u_t pa pritok novih naložb v dano delnico, ki se meri ob koncu obdobja
- Lahko optimiziramo glede na končen ali neskončen čas v prihodnosti
- Funkcija koristnosti U je podana kot

$$\sum_{t=0}^{T} \beta^t f(u_t, x_t), \tag{1}$$

  kjer $0 < \beta < 1, \beta^0 = 1, \lim_{t\to\infty} \beta^t = 0$
- β predstavlja časovno diskontiranje oziroma nestrpnost.
- To in še nekaj ostalih privzetkov nam da t.i. *problem optimalne kontrole*.

Optimalna kontrola in dinamično programiranje

Definicija (Problem optimalne kontrole)

Problem optimalne kontrole (OCP) *je formuliran kot:*

Najdi $\{u_t^*, x_t\}_{t=0}^{T}$ *(najpreprostejši OCP) oziroma najdi* $\{u_t^*, x_t\}_{t=0}^{T-1}$ *(OCP s prostim končnim stanjem), ki reši*

$$\max_{\{u_t\}_{t=0}^{T}} \sum_{t=0}^{T} \beta^t f(u_t, x_t), \tag{2}$$

kjer $u_t \in \mathcal{U}$ *in* $x_{t+1} = g(x_t, u_t)$

za podana x_0 *in* x_T *in poljuben* T*.*

Dosegljivi kandidati za optimum bojo vse poti $\{x_t, u_t\}$, ki zadostujejo pogoju $x_{t+1} = g(x_t, u_t)$ ob podanem x_0 in kateri koli izbiri $u_t \in \mathcal{U}$ v vsakem obdobju $t = \{0, \ldots, T\}$.

Dinamično programiranje

Po Bellmanovem načelu dinamičnega programiranja mora imeti optimalna pot lastnost, da morajo biti za katero koli začetno vrednost stanja x_0 in vrednost stanja x_t ter kontrole u_t na začetku obdobja t, spremenljivke kontrole izbrane optimalno za preostala obdobja, kjer smo vrednost stanja v trenutnem obdobju dobili na podlagi predhodnih optimalnih odločitev.

Dinamično programiranje

Definicija (Vrednostna funkcija)

Vrednostna funkcija *za čas τ je:*

$$V_{T-\tau}(x_\tau) = \sum_{t=\tau}^{T} \beta^{t-\tau} f(u_t^*, x_t^*) \tag{3}$$

Trditev

Optimalna rešitev problema optimalne kontrole $\{u^*, x_t^*\}_{t=0}^{T}$ zadostuje Hamilton-Jacobijevi enačbi

$$V_{T-t}(x_t) = \max_{u_t}\{f(x_t, u_t) + \beta V_{T-t-1}(x_{t+1})\} \tag{4}$$

Dinamično programiranje

Maksimizacije se lotimo v dveh korakih.

- Najprej je potrebno z rekurzijo poiskati zaporedje $\{V_T, \ldots, V_0\}$:

$$V_{t+1}(x) = \max_u \{f(x,u) + \beta V_t(g(x,u))\} \tag{5}$$

- Na podlagi dobljenega z reševanjem HJB enačbe dobimo optimalno kontrolo.

V primeru neskončnega časa, je $V = \lim_{j \to \infty} V_j$ neodvisen od j, zato HJB enačba postane:

$$V(x) = \max_u \{f(x,u) + \beta V[g(x,u)]\} = \max_u H(x,u) \tag{6}$$

Dinamično programiranje

Ob določenih predpostavkah lahko optimalno kontrolo dobimo preko stacionarne točke:

$$\frac{\partial H(x,u)}{\partial u} = 0 \tag{7}$$

Če je $H \in C^2$, dobimo funkcijo odločanja:

$$u^* = h(x), \tag{8}$$

ki nam določa optimalno pravilo za spreminjanje optimalne kontrole glede na stanje ekonomije.

Dinamično programiranje

V primeru, da tako razmerje obstaja, lahko rečemo, da je naš problem rekurziven in HJB enačba postane sledeča:

$$V(x) = f(x, h(x)) + \beta V[g(x, h(x))], \tag{9}$$

od kjer se vrednosti $V(x)$ običajno išče numerično, saj je eksaktno računanje možno le zelo poredkoma.

Primer: The cake eating problem

- Imamo torto, katere velikost v času t je označena z W_t in požrešneža, ki bi rad jedel v T obdobjih.
- $W_0 = \phi$ in $W_T = 0$
- Požrešnež ima podan psihološki diskontni faktor $0 < \beta < 1$ in logaritemsko funkcijo koristnosti.
- Želimo ugotoviti, s kakšno strategijo bo jedel torto, da bo dosegel kar se da veliko korist.

Primer: The cake eating problem

Naš cilj je torej najti optimalne poti $C^* = \{C_t^*\}_{t=0}^{T-1}$ in $W^* = \{W_t^*\}_{t=0}^{T}$, ki rešijo problem:

$$\max_{C} \sum_{t=0}^{T} \beta^t \ln(C_t), \tag{10}$$

kjer imamo določeno odvisnost $W_{t+1} = W_t - C_t$ in pogoja $W_0 = \phi$ in $W_T = 0$.

(v tej notaciji so u od prej označeni s C, x pa z W).

Primer: The cake eating problem

Da lahko rešimo ta primer z dinamičnim programiranjem, se poslužimo HJB kot v zadnji trditvi:

$$V_{T-t}(W_t) = \max_{C_t}\{ln(C_t) + \beta V_{T-t-1}(W_{t+1})\}, \quad t = 0, 1, \dots, T-1 \quad (11)$$

Upoštevati moramo še rekurzivno zvezo za preostanek torte v obdobju $t+1$, ki je enaka $W_{t+1} = W_t - C_t$.

Primer: The cake eating problem

Optimalno funkcijo odločanja za količino zaužite torte dobimo iz

$$\frac{\partial}{\partial C_t}(\ln(C_t) + \beta V_{T-t-1}(W_{t+1})) = 0, \tag{12}$$

kar nam da

$$C_t^* = C_t(W_{t+1})) = (\beta V'_{T-t-1}(W_{t+1}))^{-1} \tag{13}$$

HJB postane parcialna diferencialna enačba:

$$V_{T-t}(W_t) = ln(C_t(W_{T+1})) + \beta V_{T-t-1}(W_{t+1}), \quad t = 0, 1, \ldots, T-1 \tag{14}$$

Primer: The cake eating problem

Ko vse skupaj rešimo, v končni fazi dobimo:

$$C_t^* = \left(\frac{1 - \beta}{1 - \beta^{T-t}} \right) W_t. \tag{15}$$

Ideja vpeljave na stohastično dinamično programiranje

Spremenljivke x_t in u_t postanejo slučajne spremenljivke, iz tega sledi tudi, da so rekurzivna za x_{t+1}, koristnostna in ostale funkcije vse take, da za argumente vzamejo slučajne spremenljivke, zato je vse skupaj nekoliko težje in zahteva nekaj več znanja slučajnih procesov.