PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA
Departamento de Ingeniería Industrial y de Sistemas
Departamento de Ingeniería de Transporte y Logística

Curso: ICS3105
Semestre: 02-2021
Profesor: Mathias A. Klapp

Tarea 3 - Optimización Dinámica

Fecha de entrega: 7 de Noviembre 2021 : 23:59 HRS

Instrucciones: Se recomienda usar Latex. Evite respuestas largas y sólo responda lo que se pregunta. El corrector puede solicitar revisar su código fuente, pero su respuesta debe ir completamente en el reporte. Se prohíbe discutir la tarea fuera del grupo.

Problema 1: Algo de Teoría (20 puntos)

Considere un MDP de maximización con horizonte infinito descontado por $\lambda \in [0, 1)$ con espacio de estados $\mathbb{S}$ finito; espacio de decisiones $\mathbb{X}(s)$ finito para cada $s \in \mathbb{S}$; probabilidad de transición $p(s'|s, x)$; y función de valor inmediato $r(s, x)$ finito. Considere el valor óptimo $V^*(s)$ para cada $s \in \mathbb{S}$, donde $\mathbf{V}^*$ es su vector.

1. Considere política estacionaria cualquiera definida por la regla de decisión $d' : \mathbb{S} \to \mathbb{X}$, donde su matriz de transición de estados $p(j|i, d(i))$ para cada par de estados $i, j \in \mathbb{S}$ se denota como $\mathbf{P}_{d'}$ y su retorno inmediato $r(i, d(i))$ para cada estado $i \in \mathbb{S}$ se denota como $\mathbf{r}_{d'}$.

   El vector de valor esperado futuro descontado asociado a la política d' se define igual a

   $$\mathbf{V}^{d'} = \sum_{t=1}^{\infty} \lambda^{t-1} \cdot \mathbf{P}_{d'}^{t-1} \cdot \mathbf{r}_{d'}$$

   (a) Desde la definición de $\mathbf{V}^{d'}$ derive la ecuación de recursión $\mathbf{V}^{d'} = \mathbf{r}_{d'} + \lambda \cdot \mathbf{P}_{d'} \cdot \mathbf{V}^{d'}$ y luego obtenga una fórmula exacta para $\mathbf{V}^{d'}$.

   (b) Defina la función vectorial $L_{d'} : \mathbb{R}^{|\mathbb{S}|} \to \mathbb{R}^{|\mathbb{S}|}$ como:

   $$L_{d'}(\mathbf{V}) = \mathbf{r}_{d'} + \lambda \cdot \mathbf{P}_{d'} \cdot \mathbf{V}.$$

   Considere la serie $\mathbf{V}^{n+1} = L_{d'}(\mathbf{V}^n)$ para $n \geq 0$ considerando un punto de partida $\mathbf{V}^0$ cualquiera. Muestre que si para algún $n' \geq 0$ se cumple que $\mathbf{V}^{n'+1} \geq \mathbf{V}^{n'}$, entonces se cumple que $\mathbf{V}^{m+1} \geq \mathbf{V}^m, \quad \forall m \geq n'$.

2. Defina la función vectorial $L : \mathbb{R}^{|\mathbb{S}|} \to \mathbb{R}^{|\mathbb{S}|}$ para cada $s \in \mathbb{S}$ como:

   $$L(\mathbf{V})(s) = \max_{x \in \mathbb{X}(s)} \left\{ r(s, x) + \lambda \cdot \sum_{s' \in \mathbb{S}} p(s'|s, x) V(s') \right\}.$$

   (a) Muestre que si $\mathbf{V} \leq L(\mathbf{V})$, entonces $\mathbf{V}$ es una cota inferior (componente a componente) del vector de value-to-go $\mathbf{V}^*$ que resuelve $\mathbf{V}^* = L(\mathbf{V}^*)$.

   (b) Considere que usted está ejecutando Iteración de Valor, es decir, está computando la serie $\mathbf{V}^{n+1} = L(\mathbf{V}^n)$ desde un punto de partida $\mathbf{V}^0$ cualquiera. Muestre que si en una iteración intermedia n' se cumple que $\mathbf{V}^{n'+1} \geq \mathbf{V}^{n'}$, entonces para todo $m \geq n'$ se debe cumplir que $\mathbf{V}^{n'} \leq \mathbf{V}^m \leq \mathbf{V}^*$.

   (c) Considere que usted ejecutó Iteración de Valor hasta un error de 10^{-2}, es decir, computó la serie $\mathbf{V}^{n+1} = L(\mathbf{V}^n)$ desde un punto de partida $\mathbf{V}^0 = \mathbf{0}$ hasta una iteración n tal que $||\mathbf{V}^{n+1} - \mathbf{V}^n||_\infty = 10^{-2}$ (norma infinito). Su regla de decisión es $d'(s) = \text{argmax}_{x \in \mathbb{X}(s)} \{r(s, x) + \lambda \sum_{s' \in \mathbb{S}} p(s'|s, x) V^{n+1}(s')\}$ para todo $s \in \mathbb{S}$.

   Si $\lambda = 0,95$ ¿A qué garantía de optimalidad se encuentra d' de la política óptima?

Problema 2: Estudio óptimo (13 puntos)

Un alumno desea maximizar el uso de su tiempo y decidir una política de estudio cada semana. Suponga que el alumno puede estar en tres estados: descansado (0), estresado (1) y colapsado (2).

Si el alumno está en los estados descansado o estresado, entonces puede estudiar poco ($x = 0$) y obtener un beneficio semanal de 1 o estudiar mucho ($x = 1$) y obtener un beneficio semanal de 2. Un alumno colapsado no tiene decisiones y su beneficio semanal es 0 hasta terminar el año (a perpetuidad).

Si el alumno está descansado y estudia poco, entonces estará descansado para la próxima semana. Por el contrario, si está descansado y estudia mucho existe un 50% de probabilidades de estresarse para la próxima semana, de lo contrario termina descansado.

Si el alumno está estresado y estudia poco, hay un 50% de probabilidad de terminar descansado para la próxima semana, de lo contrario termina estresado. Si estudia mucho al estar estresado, tiene un 50% de probabilidad de terminar colapsado para la próxima semana, de lo contrario termina estresado.

Preguntas:

1. Modele este problema como un MDP de horizonte infinito descontado con $\lambda = 4/5$. Específicamente, defina su función de value-to-go y las ecuaciones de optimalidad de Bellman.
2. Escriba las ecuaciones que determinan el valor futuro esperado en cada estado de la política "estudiar poco independiente del estado". Resuélvalas, obteniendo el valor de dicha política en cada estado.
3. Resuelva el problema a optimalidad mediante Iteración de Política desde la política "estudiar poco independiente del estado".
4. Resuelva el problema a optimalidad mediante Iteración de Valor desde value-to-go nulos. Compare esta solución con la anterior en términos de resultado y cómputo.

Problema 3: Robot (13 puntos)

Supongamos que usted debe darle instrucciones a un robot saltarín que debe subir por los peldaños de una escalera de 100 pisos. El robot saltarín comienza en el piso 1 y su meta es llegar al piso 100. En cada piso s usted debe definir la acción del robot entre cuatro disponibles definidas por un parámetro $p \in \left\{\frac{1}{10}; \frac{1}{5}; \frac{2}{5}; 1\right\}$. Si se escoge la acción p, el robot saltara un piso (terminará en el piso $s+1$) con probabilidad p, saltará 2 con probabilidad $p \times (1-p)$ y así sucesivamente (saltará q pisos con probabilidad $p \cdot (1-p)^{q-1}$). El problema es que si salta más allá del piso 100 caerá al piso 1 y comenzará desde cero. Suponga que el costo de cada movida es \$1 y que el factor de descuento es $\lambda = 0,99$.

1. Modele el problema como un MDP de horizonte infinito.
2. Encuentre una política óptima que minimice el costo total esperado de subir al piso 100.

Problema 4: Inventario Estocástico (14 puntos)

Considere un problema de inventario estocástico como el desarrollado en clase, donde un tomador de decisiones (*i.e.*, usted) debe ordenar productos al comienzo de cada semana (después de observar inventario) a lo largo de un horizonte de ejecución infinito. Suponga una tasa de descuento semanal λ. Específicamente, suponga que inicia en la semana 1 con $y_1 = 30$ unidades almacenadas y que su bodega posee capacidad para $Q = 200$ unidades. La dinámica de eventos en cada semana $t \in \{1, \ldots, \infty\}$ es la siguiente:

1. Usted observa el inventario del producto en bodega $y_t \in \{0, 1, 2, \ldots, Q\}$.
2. Luego, decide ordenar $x_t \in \{0, 1, \ldots, Q - y_t\}$ unidades de producto a un costo de ordenamiento $\$1600 + 20 \cdot x_t$. El inventario ordenado es repuesto instantáneamente hasta $y'_t = y_t + x_t$.
3. Observa la demanda D_t de la semana t, donde D_t es una variable aleatoria independiente (i.i.d. para cada semana). El cliente recibe todo el producto demandado disponible (es decir, $\min(D_t, y'_t)$) y el sobrante se almacena para la siguiente semana. El costo por unidad-semana en inventario es $h = \$4$ (a pagar si un producto pasa en inventario de una semana a la otra) y el costo por unidad demandada no satisfecha es $q = \$75$. Asuma que la demanda semanal D_t es discreta y distribuida uniforme en $\{30, \ldots, 60\}$.

Preguntas:

- Modele el problema que planifica una política de decisión que minimiza el costo total esperado como un Proceso de Decisión Markoviana (MDP) de horizonte infinito.
- Obtenga una política estacionaria d que minimiza el costo total descontado esperado del sistema ejecutando iteración de valor para $\lambda \in \{90\%, 95\%, 99\%\}$ con 10^{-4} unidades de precisión absoluta, *i.e.*, $\max_{s \in \mathbb{S}}(V^d(s) - V^*(s)) < 10^{-4}$. Verifique (en cada caso) que efectivamente la política óptima obtenida posee una estructura (s, S) para cada semana.
- Optimice el problema de nuevo, pero ejecutando iteración de política para cada tasa de descuento (comience desde la política $(0, 0)$). Explote en su cómputo que la política óptima posee una estructura (s, S).
- Optimice el problema otra vez, pero ejecutando el método de programación lineal (dual o primal, usted elige cual). Confirme el resultado anterior y compare la eficiencia de cómputo de los tres métodos aplicados a este problema.
- Ahora supongan que cada vez que ordena productos existe un *Lead Time* L_t, es decir, un tiempo de entrega de la orden que tarda una semana con probabilidad 50%, dos semanas con probabilidad 35% y tres semanas con probabilidad 15%. Modele el problema de inventario estocástico de horizonte infinito con esa nueva consideración y optimice el problema que minimiza el costo total esperado descontado con el método que usted prefiera. Note que en cada período puede pedir, incluso si el pedido anterior no ha llegado.