2.098/6.255/15.093 - Recitation 8

Michael Frankovich and Shubham Gupta

November 13, 2009

1 Dynamic Programming

The number of crimes in 3 areas of a city as a function of the number of police patrol cars assigned there is indicated in the following table:

n	0	1	2	3
Area 1	14	10	7	4
Area 2	25	19	16	14
Area 3	20	14	11	8

We have a total of only 3 police cars to assign. Solve the problem of minimizing the total number of crimes in the city by assigning patrol cars using dynamic programming.

Solution.
Firstly, we define the following elements of our dynamic program. We let $N = 3$, so we have 4 stages. At $k = 0$, we assign some number of cars to area 1, then we are finished with area 1. Then at stage $k = 1$ we assign from our remaining cars some number to area 2, and so on. At $k = N = 3$, we are done and any leftover cars have no cost.

1. State x_k = number of patrol cars available at stage k;
2. Control u_k = number of patrol cars to assign at stage k to area $k + 1$;
3. Randomness ω_k constant;
4. Dynamics: $x_{k+1} = x_k - u_k$;
5. Boundary Conditions: $J_N(x_N) = 0, \quad \forall x_N$;
6. Recursion: $J_k(x_k) = \min_{u_k \in \mathcal{U}_k} \left[g_k(x_k, u_k, \omega_k) + J_k(x_{k+1})\right] = \min_{u_k \in \mathcal{U}_k} \left[g_k(x_k, u_k) + J_k(x_k - u_k)\right]$.

$$
\begin{aligned}
J_2(x_2) &= \min_{u_2 \in \{0,\dots,x_2\}} \big[g_2(x_2,u_2) + 0\big] \\
\implies J_2()^\top &= [20, 14, 11, 8] \text{ (ie notation for } J_2(0) = 20, J_2(1) = 14, \text{ etc).}
\end{aligned}
$$

$$
\begin{aligned}
J_1(x_1) &= \min_{u_1 \in \{0,\dots,x_1\}} \big[g_1(x_1,u_1) + J_2(x_1 - u_1)\big] \\
\implies J_1()^\top &= [25+20, \min\{25+14, 19+20\}, \min\{25+11, 19+14, 16+20\}, \\
&\qquad \min\{25+8, 19+11, 16+14, 14+20\}] \\
&= [45, 39, 33, 30].
\end{aligned}
$$

$$
\begin{aligned}
J_0(3) &= \min_{u_0 \in \{0,\dots,3\}} \big[g_0(x_0,u_0) + J_1(x_0 - u_0)\big] \\
&= \min\{14+30, 10+33, 7+39, 4+45\} = 43.
\end{aligned}
$$

So the optimal cost is 43 crimes. Tracing the argminima, we see that the optimal solution is to assign one car to each of the three areas.

2 Linear Algebra/Calculus Review for NLP

Definition. *A* norm $\|\cdot\|$ *on* $\mathbb{R}^n$ *is a mapping from* $\mathbb{R}^n$ *to* $\mathbb{R}$ *that satisfies:*

a) $\|x\| \geq 0, \quad \forall x \in \mathbb{R}^n$,

b) $\|cx\| = |c| \cdot \|x\|, \quad \forall c \in \mathbb{R},\ \forall x \in \mathbb{R}^n$,

c) $\|x\| = 0 \iff x = 0$,

d) $\|x+y\| \leq \|x\| + \|y\|, \quad \forall x, y \in \mathbb{R}^n$.

The following are common norms:

- The *Euclidean Norm* (or L_2-norm): $\|x\|_2 = \sqrt{x^\top x} = \left(\sum_{i=1}^n x_i^2\right)^{\frac{1}{2}}$;
- The L_1-norm: $\|x\|_1 = \sum_{i=1}^n |x_i|$;
- The p-norm ($p \geq 1$): $\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}$ (L_1 and L_2 are p-norms);
- The L_∞-norm (or max norm): $\|x\|_\infty = \max\big\{|x_1|, \dots, |x_n|\big\}$.

Let A be a real-valued symmetric (i.e. $A = A^\top$) $n \times n$ matrix. Then:

- Its eigenvalues are real.
- The following are equivalent:

a) A is positive definite.
b) All eigenvalues of A are > 0.
c) $x^\top Ax > 0, \quad \forall x \in \mathbb{R}^n \setminus \{0\}$.

- The following are equivalent:

a) A is positive semi-definite.
b) All eigenvalues of A are ≥ 0.
c) $x^\top Ax \geq 0, \quad \forall x \in \mathbb{R}^n$.

Definition. *Let $f : \mathbb{R}^n \to \mathbb{R}$. Then, when they exist,*

- $\frac{\partial f}{\partial x_i} = lim_{\alpha \to 0} \frac{f(x+\alpha e_i) - f(x)}{\alpha}$ *is the* i^{th} partial derivative *of* f *at* x.
- $\bigtriangledown f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$ *is the* gradient *of* f *at* x.
- $\bigtriangledown^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$ *is the* hessian *of* f *at* x.

3 How to determine whether a function is convex

Once we know a few basic classes of convex functions, we can use the following facts:

- Linear functions $f(x) = a^\top x + b$ are convex.
- Quadratic functions $f(x) = \frac{1}{2} x^\top Qx + b^\top x$ are convex if Q is PSD (positive semi-definite).
- Norms are convex functions (the proof is left an exercise, using the properties of norms defined above).
- $g(x) = \sum_{i=1}^{k} a_i f_i(x)$ is convex if $a_i \geq 0, f_i$ convex, $\forall i \in \{1, \ldots, k\}$.

Alternatively, if a function is differentiable, we can use the following facts:

- $\bigtriangledown^2 f(x)$ is PSD $\forall x \implies$ f is convex.
- $\bigtriangledown^2 f(x)$ is PD (positive definite) $\forall x \implies$ f is strictly convex.

Finally, if the function is not differentiable and we cannot use one of the above approaches, we check the definition of convexity:

Definition. *A function* $f : \mathbb{R}^n \to \mathbb{R}$ *is* convex *if* $\forall x, y \in \mathbb{R}^n$*, we have*

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y), \quad \forall \lambda \in [0,1].$$

3.1 Example

Let $f : \mathbb{R}^n \to \mathbb{R}, (x_1, x_2) \mapsto x_1 x_2^2 - x_1$. So

- $\bigtriangledown f(x) = \begin{bmatrix} x_2^2 - 1 \\ 2x_1 x_2 \end{bmatrix}$,
- $\bigtriangledown^2 f(x) = \begin{bmatrix} 0 & 2x_2 \\ 2x_2 & 2x_1 \end{bmatrix}$.

To solve for the eigenvalues of the hessian, we get the following quadratic in λ:

$$\det\big(\bigtriangledown^2 f(x)\big) = \det \begin{bmatrix} -\lambda & 2x_2 \\ 2x_2 & 2x_1 - \lambda \end{bmatrix} = 0,$$
$$\lambda^2 - 2x_1\lambda - 4x_2^2 = 0.$$

Since the constant term is negative, we cannot have two roots (i.e. eigenvalues) of the same sign. Hence f can be neither convex nor concave.

15.093J / 6.255J Optimization Methods
Fall 2009