Recitation Notes on MDPs and Learning Bayes Nets

Hoda Heidari

December 4, 2017

1 MDPs

In an MDP, the decision maker's reward, r, at each time step could be a function of its current state, x, the action, a, it takes at that time step, and the state in moves to as the result of that action, x'. That is

$$r = r(x, a, x').$$

The above is the most general case, and the formula for computing the value function $V_\pi(x)$ in this case is as follows:

$$V_\pi(x) = \sum_{x' \in X} P(x'|x, \pi(x)) \left(r(x, \pi(x), x') + \gamma V_\pi(x') \right) \tag{1}$$

We solved a problem (Vacuum Cleaning Robot, from exam 2012) in the Nov 24th recitation that used this general formula.

We also saw two special cases of the above formulation in the recitation and lecture:

- The reward is a function of current state only (i.e. $r(x, a, x') = r(x)$). The value function in this case simplifies to:

  $$V_\pi(x) = r(x) + \gamma \sum_{x' \in X} P(x'|x, \pi(x)) V_\pi(x') \tag{2}$$

  We solved a problem (Managing a shop with MDPs, from exam 2016) in the Nov 24th recitation that used this formula.

- The reward is a function of current state and action only (i.e. $r(x, a, x') = r(x, a)$). The value function in this case simplifies to:

  $$V_\pi(x) = r(x, \pi(x)) + \gamma \sum_{x' \in X} P(x'|x, \pi(x)) V_\pi(x') \tag{3}$$

Note that one can still apply equation (1) to the above two special case settings and get the exact same numerical results: (2), (3) are just the simplified/shortened version of (1) when the reward function has fewer parameters.

2 Learning Bayes Nets

The mutual information between two random variables X, Y is denoted by $I(X;Y)$, and is related to entropy, H, as follows:

$$I(X;Y) = H(X) - H(X|Y)$$

We used the relation between entropy and mutual information to solve a problem from exam 2012 (Information Gain, part (c)) in Dec 1st recitation. In case you have not seen entropy before, the entropy of a random variable X is defined as follows:

$$H(X) = \sum_x -P(X = x) \log(P(X = x)).$$

Here are some properties of $I(X, Y)$ that you should know:

- $I(X;Y) \geq 0$.
- $I(X;Y) = I(Y;X)$. We used this property to solve a problem from exam 2014 (Learning Bayesian Networks, part (a)).
- For any set $S \subseteq T$ of random variables

  $$I(X;S) \leq I(X;T)$$

  Note that a set of random variables is itself a random variable! So $I(X;S)$ and $I(X;T)$ are well-defined. We used this property to solve a problem from exam 2012 (Information Gain, part (a)).

  Also note that this property is responsible for why $S(G;D)$ is always maximized for the complete network. Remember that

  $$S(G;D) = \sum_{i=1}^{n} I(X_i, \mathbf{Pa}_i)$$

  so one can always increase the score by expanding $\mathbf{Pa}_i$.