CPSC-540 Machine Learning 2005

Homework # 1
Due Thursday Feb 23 in class.

NAME:_____

Signature:_____

STD. NUM: _____

General guidelines for homeworks:

You are encouraged to discuss the problems with others in the class, but all write-ups are to be done on your own.
Homework grades will be based not only on getting the "correct answer," but also on good writing style and clear presentation of your solution. It is your responsibility to make sure that the graders can easily follow your line of reasoning.
Try every problem. Even if you can't solve the problem, you will receive partial credit for explaining why you got stuck on a promising line of attack. More importantly, you will get valuable feedback that will help you learn the material.
Please acknowledge the people with whom you discussed the problems and what sources you used to help you solve the problem (e.g. books from the library). This won't affect your grade but is important as academic honesty.
When dealing with Matlab exercises, please attach a printout with all your code and show your results clearly.

1. **Weak Law of Large Numbers**:

   Let $X_{1:n}$ be a sequence of i.i.d. random variables with $\mathbb{E}(X_i) = \mu$ and $var(X_i) = \sigma^2$. Let also

   $$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

   (i) Markov's inequality: For a random variable X and any positive measurable function $f(X)$ and positive scalar ε, show that

   $$P(f(X) \geq \varepsilon) \leq \frac{\mathbb{E}(f(X))}{\varepsilon}$$

   Hint: express $f(X)$ as $f(X) \geq \varepsilon \mathbb{I}_{f(X)\geq\varepsilon}$ and apply the expectation operator.

   (ii) Chebyshev's inequality: Choose an appropriate $f(\overline{X}_n)$ in (i) to show:

   $$P(|\overline{X}_n - \mu| \geq \varepsilon) \leq \frac{var(\overline{X}_n)}{\varepsilon^2}$$

(iii): Show that the asymptotic estimator of the mean is unbiased; $\mathbb{E}(\overline{X}_n) = \mu$.

(iv) Using the fact that for independent X_i's we have $var(\sum X_i) = \sum var(X_i)$, show that

$$var(\overline{X}_n) = \frac{\sigma^2}{n}$$

(v) Show the following weak law of large numbers:

$$P(|\overline{X}_n - \mu| \geq \varepsilon) \underset{n\to\infty}{\longrightarrow} 0$$

This statement says that the sample mean $\overline{X}_n$ converges to the true mean *in probability* as n goes to infinity. There is another mode of convergence, called *strong convergence* or almost sure convergence, which asserts a bit more. $\overline{X}_n$ is said to converge *almost surely* to μ if for every $\varepsilon > 0$, $|\overline{X}_n - \mu| \geq \varepsilon$ happens only a finite number of times *with probability 1*. Finally, to study the speed of convergence, one introduces *central limit theorems*.

2. **Linear Programming for MDPs**: For any value function V and Bellman operator T we know that:

   (a) If $V \geq TV$, then $V \geq V^\star$.

   (b) If $V \leq TV$, then $V \leq V^\star$.

   (c) If $V = TV$, then $V = V^\star$.

   That is, applying T to V takes us to a fixed point $V^\star$. We can use this result to cast the MDP problem in terms of linear programming:

   Primal:

$$\text{Minimize} \sum_j V(j)$$

   subject to $V \geq T_d V$, that is subject to the linear system of equations:

$$V(i) \geq \sum_j p(j|i,a) \left[\gamma V(j) + r(i,a,j)\right]$$

   for all a and i.

   (i) Derive an expression for the dual linear programming problem. You might want to consult a standard book with linear programming such as the optimization book of Nocedal and Wright. Also google the topic.

.

(i) Solve the Gridworld MDP using the matlab files provided on the website and your implementation of linear programming. Hint: Look at the INRA MDP toolbox in the course's matlab website. Compare the accuracy vs computational time of policy iteration, value iteration and your linear programming algorithm (i.e. generate a value for money plot).

3. **Q-Learning and Sarsa**: Implement the Q-learning and Sarsa algorithms on the Cliffworld example of Sutton. Confirm the results of Figure 6.14.

4. **Contraction Operators:** In pages 55 to 57 of the notes, we proved that T_d is a contraction operator for the value function. Prove that T is also a contraction operator. Hint: show first that

$$TV(i) \leq T\overline{V}(i) + \beta\xi(i) \max_j \frac{\|V(j) - \overline{V}(j)\|_\xi}{\xi(j)}$$

interchange V and $\overline{V}$, and use the equation resulting from this interchange to get the right absolute value.

5. **Convergence of Policy Iteration**: Prove Theorem 4 on page 61 of the course notes. Hint: By design, given a policy d_t, policy improvement chooses d_{t+1} such that $T_{d_{t+1}} V^{d_t} = TV^{d_t}$. Also recall that the optimal cost is lowest, $V^{d_t} \geq V^\star$, and the number of valid policies is finite and less than $|\mathcal{A}|^{|\mathcal{X}|}$.