Advanced topics in Machine Learning
Assignment 5

Mariuta Nicolae (rqt629) October 18, 2015

Question 1

For this assignment, I have decided to implement the Iterative policy evaluation algorithm because that one I understood better.

I started by defining the main parameters for the maze that was given: the 12 states corresponding to each of the rooms, the 4 possible actions for movement, the parameter for discount factor γ that is 0.5 but I have experimented with different values for this parameter and the stopping criterion, the parameter θ that is set to 0.001. After that I defined a matrix where I keep the values for optimization of function V with 12 rows and 4 columns, I also created a matrix with the possible rewards for tanking each of the 4 action from each of the states and also a matrix where I define what is the next state for each action made in each room that is the transitions matrix.

I defined to have maximum 1000 loops for the algorithms or it stops if delta is smaller than the parameterθ. At each step I made the calculation as in the algorithm shown in the lecture and at each step I recalculate values in V.

At the end I take the maximum value from each row of V to see what the best action to make in each case is.

The 12 values in optimal value function V are: up, up, down, left, up, up, up, up, up, up, right, down.

I think something went wrong because in the first room the agent will hit the wall if it goes up and it does not look like I am going to reach the destination with this result. I am not sure what went wrong but I think I understood something wrong in the algorithm and I obtained these weird results.

I also did not understand how to implement using the random policy but I have used only this algorithm to obtain the optimal value function.

Question 2

1.

a) From what I understand, the matrix **a** is a matrix with only one row and n columns so its transpose will be a column matrix. If we multiply a row matrix with a column matrix, we then obtain a matrix with a single element and the rank of a matrix with 1 row and 1 column can only be 1.

b) So I have to prove that **a** is eigenvector of matrix **C** so I have to find that:

$$Ca = \lambda a$$

where λ is the scalar that is eigenvalue of C. Since a is row matrix and C is a matrix with only one element, as shown previously, then C is also a scalar so the eigenvalue of C is actually that element from C. Then, we will have something like:

$$Ca = Ca$$

which is obviously true, so we have that a is eigenvector of C and that C is the eigenvalue. Maybe the proof is not formal enough but I think this is the idea.

c) If I consider the definition of covariance, we have:

$$\sigma^2 = \frac{1}{N}\sum_i \left(x^{(i)} - m\right)^2$$

Because the mean is 0, it means we have:

$$\sigma^2 = \frac{1}{N}\sum_i \left(x^{(i)}\right)^2$$

So the covariance should be the squared sum of b that is a 1d matrix, so in order to obtain b with highest probability we must have the distribution:

$$N\left(0, \frac{1}{N}\sum_i \left(x^{(i)}\right)^2\right)$$

where $x^{(i)}$ are the elements in matrix b. I am not sure if this is correct but this is what I obtain by applying the definitions.

2.

a) For this exercise, I should basically what is the mean of sum of the vectors $x_1, \dots, x_m$ and also the covariance of this sum of the vectors.

For the mean it is quite obvious that if the mean of each vector is vector 0, then the sum of these vectors also have mean 0 since we have even for matrices that: $\frac{X+Y}{n} = \frac{X}{n} + \frac{Y}{n}$.

Regarding the covariance, there is property that the variance of sum of vectors is equal with sum of the variance for each vector:

$$Var(A + B) = Var(A) + Var(B)$$

Now I guess I should also prove this is true so I will also start from the definition and we have the covariance of x_1:

$$\sigma^2(x_1) = \frac{1}{N}\sum_i (x_1^{(i)} - 0)\text{^}2$$

If we sum the covariance of 2 vectors:

$$\sigma^2(x_1) + \sigma^2(x_2) = \frac{1}{N}\sum_i (x_1^{(i)})\text{^}2 + \frac{1}{N}\sum_i (x_2^{(i)})\text{^}2 = \frac{1}{N}\sum_i (x_1^{(i)} + (x_2^{(i)})\text{^}2 = \sigma^2(x_1 + x_2)$$

I have ignored the mean since it is 0 so I have somehow shown that:

$$\sigma^2\left(\sum_{i=1}^{m} x_i\right) = \sum_{i=1}^{m} \sigma^2(x_i) = \sum_{i=1}^{m} I = mI$$

So I obtained for $\sum_{i=1}^{m} x_i$ the distribution: $N(0, mI)$ so mean 0 and covariance diagonal of m, the number of vectors.

b) In this case, the mean will still be zero because we have the mean:

$$mean = w_1 \frac{\sum x_1}{n} + w_2 \frac{\sum x_2}{n} + \ \dots \ + w_m \frac{\sum x_m}{n} = \ w_1 0 + \ w_2 0 + \cdots + w_m 0 \ = 0$$

Now, for the covariance, it is according to all documentation the property that:

$$\sigma(aX, bX) = ab\sigma^2(X)$$

So in our case we will have:

$$\sigma^2\left(\sum_{i=1}^{m} w_i x_i\right) = \sum_{i=1}^{m} w_i \sigma^2(x_i) = \sum_{i=1}^{m} w_i I$$

So I obtain the distribution $N(0, \sum_{i=1}^{m} w_i I\ )$ so the same mean 0 and covariance with diagonal equal with sum of weights.

c) Since I have shown at 1a that the rank of multiplying a row matrix with its transpose we obtain a rank 1 matrix with 1 column and 1 row so I think the sum of these matrices will still have rank 1 but I am not sure of this and I do not know how to make a complete proof.

3.

For the equation:

$$(1-c)a + \sqrt{c(2-c)}b$$

We basically have 2 constants: $(1-c)$ and $\sqrt{c(2-c)}$ and we have to calculate first the mean for sum of 2 distributions that are multiplied by a constant, but I have already shown previously that:

$$mean\left((1-c)a + \sqrt{c(2-c)}b\right) = \ (1-c)mean(a) + \ \sqrt{c(2-c)}mean(b)$$
$$= \ (1-c)0 + \sqrt{c(2-c)}0 = 0$$

So the mean of the resulting distribution is still vector 0. Now for the covariance, it is written on Wikipedia that:

$$\sigma(aX + bY, cW + \mathrm{dV}) = ac\sigma(X, W) + a\mathrm{d}\sigma(X, V) + \ bc\sigma(Y, W) + b\mathrm{d}\sigma(Y, V)$$

and the proof for this I think it will take long time to calculate but I think it is done somewhat similar to what I have written at exercise 2c. In our case we have to calculate covariance and we will have:

$$\sigma(aX + bY, aX + bY) = \ a^2\sigma(X, X) + ab\sigma(X, Y) + ab\sigma(X, Y) + b^2\sigma(Y, Y)$$
$$= a^2\sigma^2(X) + 2ab\sigma(X, Y) + b^2\sigma^2(Y, Y)$$

In our case we have the replacements: $a \to (1-c)\ , b \to \ \sqrt{c(2-c)}, X \to a, Y \to b$

We know that the covariance of a and d are I so $\sigma(a,b) = I$ as I shown at the previous exercise so we have that:

$$\sigma^2\left((1-c)a + \sqrt{c(2-c)}b\right) = (1-c)^2 I + 2(1-c)\sqrt{c(2-c)}I + \sqrt{c(2-c)}\sqrt{c(2-c)}I$$
$$= I\left((1-c)^2 + 2(1-c)\sqrt{c(2-c)} + c(2-c)\right)$$

I do not know how to continue these calculations but I think we use the fact that $c\epsilon[0,1]$ and somehow prove that $\left((1-c)^2 + 2(1-c)\sqrt{c(2-c)} + c(2-c)\right) \sim 1$ and then we have:

$$(1-c)a + \sqrt{c(2-c)}b \sim N(0, I)$$

Question 3

In order to evaluate the performance of algorithms for the adversarial environments, I think we have to generate the data in the way that we have the worst performance for our algorithm. We basically have to calculate the regret of the algorithm in cases when the loss should be the highest. I think this it is very hard to make this analysis very accurate because in most of the situation, the adversary cannot see exactly how our algorithm is built and perhaps he usually not provide us the data that gives the worst regret each time.

Question 4

1. In comparison to Label-Efficient Forecaster, at the Hedge algorithm we had to observe all k columns at each run of the loop so we had to observe kT columns in total. Now we have to draw variable Z_t with the bias ε so there it will be ε chance to observe column t.

So for a loop t the expectation for observing t columns will be: $\frac{\varepsilon}{t}T$ and since we have t runs in total, the total expectation will be εT. I think this is enough for making this proof.

2. If we remove the part of the code were we draw the random variable Z_t in order to decide if we want to observer $l_t^1, l_t^2, \dots, l_t^K$ we will basically have the Hedge algorithm where I have shown for the previous assignment that we have the expected regret of the algorithm to satisfy:

$$E[R_t] \leq \sqrt{2TlnK}$$

So that is going to be the expected regret for $\varepsilon = 1$ but now we have to consider the adversarial setting we have and see the changes.

At some point, for the Hedge algorithm we had the calculation summary:

$$ln\frac{W_t}{W_0} \leq -\eta \sum_{t=1}^{T}\sum_{a} l_t^a P_t(\mathrm{a}) + \frac{\eta^2}{2}\sum_{t=1}^{T}\sum_{a} P_t(\mathrm{a})$$

$$\eta\sum_{t=1}^{T}\sum_{a} l_t^a\, P_t(\mathrm{a}) - \eta \min_{a}\left(\hat{L}_T(a)\right) \le \ln\left(\frac{K}{\eta}\right) \;+\frac{\eta}{2}\sum_{t=1}^{T}\sum_{a}(l_t^a)^2 P_t(\mathrm{a})$$

And I think here we have the difference since there it is the bias ε so we will have for our adversarial environment because we have $l_t^a P_t(a) = 0$ if we draw that $Z_t = 0$. I am not sure how to write that formally but I think it is something like:

$$\eta\sum_{t=1}^{T}\sum_{a}^{Z_t=1} l_t^a\, P_t(\mathrm{a}) - \eta \min_{a}\left(\hat{L}_T(a)\right) \le \ln\left(\frac{K}{\eta}\right) \;+\frac{\eta}{2}\sum_{t=1}^{T}\sum_{a}(l_t^a)^2 P_t(\mathrm{a})$$

And if we minimize with respect to η : $\eta = \sqrt{\frac{2lnk}{T}}$ we will obtain that:

$$E[R_t] \le \sqrt{\frac{2T}{\varepsilon} lnK}$$

I think I should have written more about how to make this proof but I do not know how to make it more formal but I think that is the main idea.