CS464 - Introduction to Machine Learning

Homework Assignment - 1

Berkan Ozdamar
21602353

Question 1

1.1

We are given that a basketball team has 8 matches left. The probability for that team to win a match is 0.6 and it is independent for each match. Also, turnover is defined as winning after a lose or losing after a win. The question asks us to compute the probability of that basketball team to have exactly 1 turnover for the remaining matches.

We can express the probability of win as:

$$P(W = 1) = 0.6$$

$$P(W = 0) = 1 - (0.6) = 0.4$$

Then, lets consider the turnover case. For having exactly 1 turnover, there must at least 1 win and 1 lose to change state. Also, regardless of the number of win and lose matches, all the wins and loses should be consecutive. In other words, win matches should be successive and lost matches should be successive. Therefore, the team can start with successive loses and end the season with successive wins or the team can start with successive wins and end the season with successive loses. So there are only 2 variations for how wins and loses can aligned regardless of the number of wins and loses. So we can eventually write the distribution as:

$$P(X|Turnover = 1) = \sum_{n=1}^{7} 2 * p^n * (1-p)^{8-n}$$

Here, we can write the p as $P(W = 1) = 0.6$. Also summation is from 1 to 7 since there must at least 1 win or 1 lose.

$$P(X|Turnover = 1) = \sum_{n=1}^{7} 2 * (0.6)^n * (0.4)^{8-n}$$

$$P(X|Turnover = 1) = \sum_{n=1}^{7} 2 * (0.6)^n * (0.4)^8 * (0.4)^{-n}$$

Here we can take the $2 * (0.4)^8$ out of the summation and update $(0.6)^n * (0.4)^- n$ as $(\frac{3}{2})^n$.

$$P(X|Turnover = 1) = 2 * (0.4)^8 * \sum_{n=1}^{7} (\frac{3}{2})^n$$

We can calculate the summation part as:

$$\sum_{n=1}^{7} (\frac{3}{2})^n = \frac{6177}{128}$$

Finally, we can write the $P(X|Turnover = 1)$ as:

$$P(X|Turnover = 1) = 2 * (0.4)^8 * \frac{6177}{128} = 0.063$$

1.2

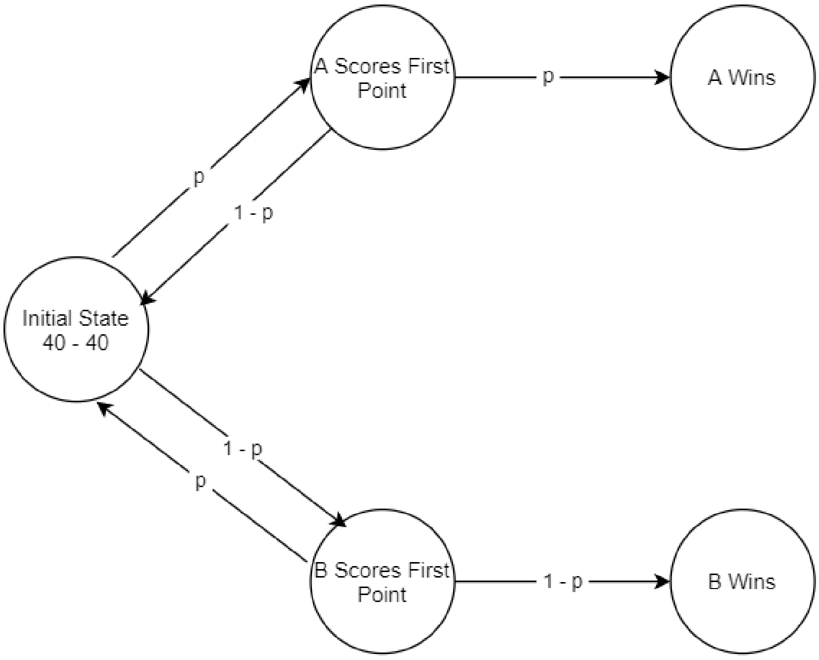
In this question we are given that the probability of player A is scoring in a tennis match is p and it is independent of the score. Also, if the score is 40-40, then 2 successive points should be made to win the match. We are asked the probability of the player A winning the game if the score is 40-40. Also we are asked to plot this probability as a function of p for p ∈ [0,1].

First, we can write the probability of player A scoring and probability of player B scoring as:

$$P(A\ scores) = p$$

$$P(B\,scores) = 1 - P(A\,scores) = 1 - p$$

Here, we can explain the states of the game with Markov Decision Process. Since we are given the current score is 40-40, initial state will be 40-40. Then, if A scores first and B scores after, the score is again 40-40 and we are back to initial state. Similarly, if B scores first and A scores after, we are again back to initial state. If A or B scores 2 successive points, then the game is over. The described Markov Decision Process can be shown as:



Here, we care about the cases that A wins. From the Markov Decision Process, we can write few scoring cases as:

AA, ABAA, BAAA, ABABAA, ABBAAA, BAABAA, BABAAA, ...

As can be seen from the exaples that, we need AA after the initial state to make A win. But the combinations of AB and BA returns the initial state. If we write the probabilities of A winning for the examples above:

$$AA:\ p^2$$

$$ABAA\ +\ BAAA:\ 2p(1-p)p^2$$

$$ABABAA\ +\ ABBAAA\ +\ BAABAA\ +\ BABAAA:\ 2^2p^2(1-p)^2p^2$$

From these examples, it can be interpreted that the last p^2 is mutual for all examples since it comes from the AA which make A win. Also $p^k(1-n)^k$ is also mutual for all these examples with different k values, since $p^k(1-n)^k$ comes from the combinations of AB and BA. Thus, we can write the probability of the player A winning the game if the score is 40-40 as:

$$P(A\,wins|Score = 40-40) = p^2 * \sum_{n=0}^{\infty} 2^n * p^n * (1-p)^n$$

Also, we know that:

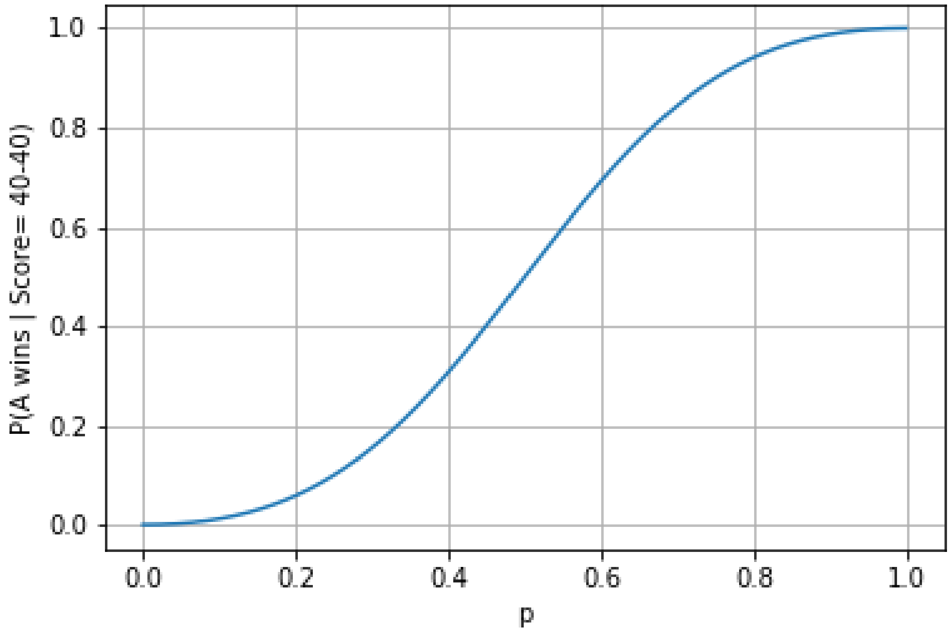
$$\sum_{n=0}^{\infty} a^n = \frac{1}{1-a}$$

where $a^n = 2^n * p^n * (1-p)^n = (2p(1-p))^n$ so the equation becomes:

$$P(A\,wins|Score = 40-40) = p^2 * \frac{1}{1-(2p(1-p))}$$

$$P(A\,wins|Score = 40-40) = \frac{p^2}{2p^2-2p+1}$$

The plot of $P(A\,wins|Score = 40-40)$ for p $\in$ [0,1] is:



Question 4

We are given that the variance z_1 which is the projection of feature set X on w_1 vector. Also;

$$Var(z_1) \;=\; w_1^T \Sigma w_1, \quad where \quad \Sigma \;=\; Cov(X)$$

Moreover, the constraint $||w_1|| = 1$ is also given.

4.1

We are required to show that the first principal component of PCA is the eigenvector of the Cov(X) with the largest eigenvalue.

We need to maximize the variance of z_1 which is:

$$Var(z_1) \;=\; w_1^T \Sigma w_1$$

where w_1 is the eigenvector for Cov(X) with the constraint $||w_1|| = 1$. To deal with the constraint while maximizing Var(z_1), we will use Lagrange multiplier. So the equation becomes;

$$maximize \quad Var(z_1) \;=\; w_1^T \Sigma w_1 \;-\; \lambda_1 (w_1^T w_1 \;-\; 1)$$

Taking derivative with respect to w_1 and setting it to zero, we get

$$\Sigma w_1 \;=\; \lambda_1 w_1$$

Multiply both sides with w_1^T

$$w_1^T \Sigma w_1 \;=\; \lambda_1 w_1^T w_1$$

$$w_1^T \Sigma w_1 \;=\; Var(z_1) \;=\lambda_1$$

Hence, eigenvector corresponding to largest eigenvalue of Cov(X) is the first principal component of PCA.

4.2

We are required to show that second principal component corresponds to eigenvector with the second largest eigenvalue.

We need to maximize the variance of z_2 which is:

$$Var(z_2) \;=\; w_2^T \Sigma w_2$$

with the constraint $||w_1|| = 1$. Also, we know that all eigenvectors are orthogonal which corresponds to $w_2^T w_1 = 0$. Taking consideration for the both constraints, we will have 2 Lagrance multiplier. The equation becomes:

$$maximize \quad Var(z_2) \;=\; w_2^T \Sigma w_2 \;-\; \lambda_1 (w_2^T w_2 \;-\; 1) \;-\; \lambda_2 (w_2^T w_1 \;-\; 0)$$

Taking derivative with respect to w_1 and setting it to zero, we get

$$\Sigma w_2 = \; \lambda_1 w_2 - \lambda_2 w_1$$

Multiply both sides with w_2^T

$$w_2^T \Sigma w_2 = \; \lambda_1 w_2^T w_2 - \lambda_2 w_2^T w_1$$

Here, $\lambda_2 w_2^T w_1 = 0$ which can be seen from the constraints. So;

$$w_2^T \Sigma w_2 = \; Var(z_2) \; = \lambda_1$$

Hence, eigenvector corresponding to second largest eigenvalue of Cov(X) is the second principal component of PCA.

References

References

[1] *Tim Roughgarden & Gregory Valiant*, "CS168: The Modern Algorithmic Toolbox Lecture 8: How PCA Works" Available: https://web.stanford.edu/class/cs168/l/l8.pdf