ITEC201006 - Introduction to Computer Science & Engineering, Jeonghun Park

HW3 – No due date

Q1. Calculate the entropy of the following distribution:

P[X = a] = 0.1

P[X = b] = 0.2

P[X = c] = 0.3

P[X = d] = 0.4

Q2. Assume that we have a set of probability **p** = [p1, p2, p3, …, pN]. Calculate the entropy of this set of probabilities, and find the probability set that makes the entropy minimum. Also find the probability set that makes the entropy maximum.

Q3. What is the entropy of the egg-drop experiment?

Q4. Consider the following table:

X \ Y	0	1
0	$\frac{1}{3}$	$\frac{1}{3}$
1	0	$\frac{1}{3}$

Find the following:

1) H(X)
2) H(Y)
3) H(X,Y)
4) I(X;Y) = H(Y) – H(Y|X)

ITEC201006 - Introduction to Computer Science & Engineering, Jeonghun Park

HW3 – No due date

Q5. Machine learning has 3 key elements, E (experience), P (Performance), and T (Task). Write the proper one to the followings:

- The real driving data for automotive driving
- The car crash rate
- Algorithm to drive a car in heavy traffic

Q6. Why gradient descent stops at the optimal point? Explain.

Q7. Assume that we have the following data. Obtain the results of the feature scaling.

X = -100 ~ 2000

Y = 0 ~ 1000

Z = -2000 ~ 100

Q8. Assume that we have only one training data (x,y) = (1,2). Perform linear regression.

Q9. (Difficult, but simple) An urn contains r red, w white, and b black balls. Which has higher entropy, drawing $k \geq 2$ balls from the urn with replacement or without replacement? Set it up and show why.

ITEC201006 - Introduction to Computer Science & Engineering, Jeonghun Park

HW3 – No due date

Q10. (Difficult) The World Series is a seven-game series that terminates as soon as either team wins four games. Let X be the random variable that represents the outcome of a World Series between teams A and B; possible values of X are AAAA, BABABAB, and BBBAAAA. Let Y be the number of games played, which ranges from 4 to 7. Assuming that A and B are equally matched and that the games are independent, calculate H(X).