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

Sol) $\mathrm{H(X)} = 0.1 \log_2 0.1^{-1} + 0.2 \log_2 0.2^{-1} + 0.3 \log_2 0.3^{-1} + 0.4 \log_2 0.4^{-1}$

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.

Sol) $\mathrm{H(X)} = \sum_{i=1}^{N} pi \log_2 pi^{-1}$.

Minimum: p1 = 1, the others are zero

Maximum: p1 = p2 = … = pN

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

Sol) $\mathrm{H(X)} = log2(N)$

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)

Sol)

$H(X) = \frac{2}{3}\log\frac{3}{2} + \frac{1}{3}\log 3 = 0.918 \text{ bits } = H(Y)$

$H(X,Y) = 3 \times \frac{1}{3}\log 3 = 1.585 \text{ bits.}$

$I(X;Y) = H(Y) - H(Y|X) = 0.251 \text{ bits.}$

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 **E**
- The car crash rate **P**
- Algorithm to drive a car in heavy traffic **T**

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

Sol) Since $\frac{\partial J(a,\cdot)}{\partial a} = 0$

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

X = -100 ~ 2000 -

Y = 0 ~ 1000

Z = -2000 ~ 100

ITEC201006 - Introduction to Computer Science & Engineering, Jeonghun Park

HW3 – No due date

Sol) Xscaled = $(X - 1050)/2100$

Yscaled = (Y – 500)/100

Zscaled = (Z + 950)/2100

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

Sol) $J(a) = (a*1 - 2)^2$, so the cost function is minimized at a = 2. h(x) = ax, a = 2.

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.

Sol) $H(X|X_{\text{prev}}) \leq H(X)$

(Try to think about why)

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).

Sol)

There are 2 (AAAA, BBBB) World Series with 4 games. Each happens with probability $(1/2)^4$.
There are $8 = 2\binom{4}{3}$ World Series with 5 games. Each happens with probability $(1/2)^5$.
There are $20 = 2\binom{5}{3}$ World Series with 6 games. Each happens with probability $(1/2)^6$.
There are $40 = 2\binom{6}{3}$ World Series with 7 games. Each happens with probability $(1/2)^7$.

The probability of a 4 game series ($Y = 4$) is $2(1/2)^4 = 1/8$.
The probability of a 5 game series ($Y = 5$) is $8(1/2)^5 = 1/4$.
The probability of a 6 game series ($Y = 6$) is $20(1/2)^6 = 5/16$.
The probability of a 7 game series ($Y = 7$) is $40(1/2)^7 = 5/16$.

$$
\begin{aligned}
H(X) &= \sum p(x) log \frac{1}{p(x)} \\
&= 2(1/16)\log 16 + 8(1/32)\log 32 + 20(1/64)\log 64 + 40(1/128)\log 128 \\
&= 5.8125
\end{aligned}
$$