# COMP9444 Neural Networks and Deep Learning Term 3, 2020

## Solutions to Exercises 3: Probability

**This page was last updated: 09/30/2020 05:13:27**

---

1. **Bayes' Rule**

   One bag contains 2 red balls and 3 white balls. Another bag contains 3 red balls and 2 green balls. One of these bags is chosen at random, and two balls are drawn randomly from that bag, without replacement. Both of the balls turn out to be red. What is the probability that the first bag is the one that was chosen?

   Let B = first bag is chosen, R = both balls are red. Then
   P ( R | B ) = (2/5)*(1/4) = 1/10
   P ( R | ¬B ) = (3/5)*(2/4) = 3/10
   P ( R )      = (1/2)*(1/10) + (1/2)*(3/10) = 1/5
   P ( B | R )  = P(R|B)*P(B) / P(R) = (1/10)*(1/2)/(1/5) = 1/4
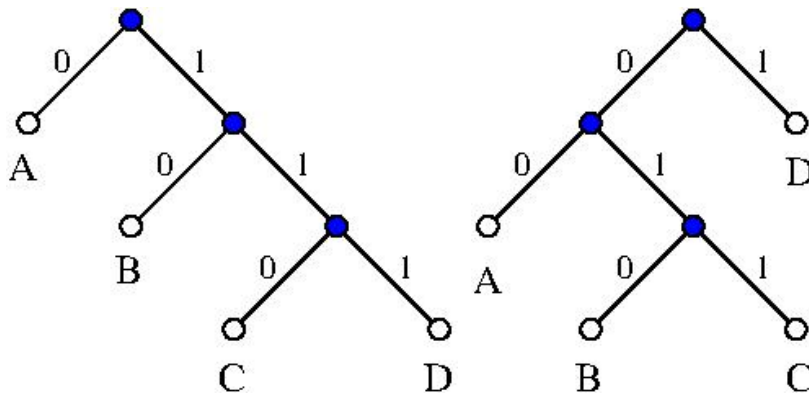
2. **Entropy and Kullback–Leibler Divergence**

   Consider these two probability distributions on the same space $\Omega$ = {A, B, C, D}

   $p = \langle\ ½, ¼, ⅛, ⅛\ \rangle$
   $q = \langle\ ¼, ⅛, ⅛, ½\ \rangle$

   a. Construct a Huffmann tree for each distribution $p$ and $q$

b. Compute the entropy H($p$)

$$H(p) = H(q) = \tfrac{1}{2}(-\log \tfrac{1}{2}) + \tfrac{1}{4}(-\log \tfrac{1}{4}) + \tfrac{1}{8}(-\log \tfrac{1}{8}) + \tfrac{1}{8}(-\log \tfrac{1}{8})$$
$$= \tfrac{1}{2}(1) + \tfrac{1}{4}(2) + \tfrac{1}{8}(3) + \tfrac{1}{8}(3) = 1.75$$

c. Compute the KL–Divergence in each direction $D_{KL}(p \parallel q)$ and $D_{KL}(q \parallel p)$

$$D_{KL}(p \parallel q) = \tfrac{1}{2}(2-1) + \tfrac{1}{4}(3-2) + \tfrac{1}{8}(3-3) + \tfrac{1}{8}(1-3) = 0.5$$

$$D_{KL}(q \parallel p) = \tfrac{1}{4}(1-2) + \tfrac{1}{8}(2-3) + \tfrac{1}{8}(3-3) + \tfrac{1}{2}(3-1) = 0.625$$

Which one is larger? Why?

$D_{KL}(q \parallel p)$ is larger, mainly because the frequency of D has increased from ⅛ to ½, so it incurs a cost of 3−1=2 additional bits every time it occurs (which is often).