AI - CSE 643 Quiz-3 Solutions, Monsoon 2023 Nov. 16, 2023

Max score: 7 Time: 40 minutes

1. (3 points) Given the Bayesian Network in Fig. 1, compute the probability by exact inference $P(Rain|Sprinkler = true)$.

$$P(R|s=t) = \alpha \sum_{c,w} P(R,s,c,w)$$

t = true , f = false

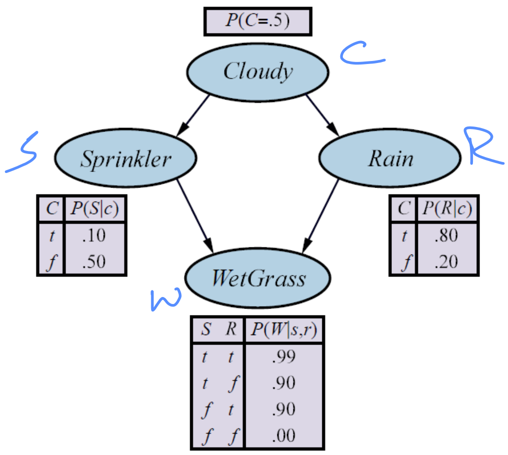


Figure 1: Bayesian Network for Q1

$$P(R=r|s=t) = \alpha \sum_{c,w} P(c)P(s|c)P(r|c)P(w|s,r)$$

for $r=t$, $s=t$

$$P(R=r|s=t) = \alpha \sum_{c} P(c)P(s|c)P(r|c) \sum_{w} P(w|s,r)$$

$$= \alpha\Big(\big[0.5 \times 0.1 \times 0.8\,(0.99+0.01)\big] + \big[0.5 \times 0.5 \times 0.2\,(0.99+0.01)\big]\Big)$$

$$= \alpha\,(0.04 + 0.05) = \alpha \times 0.09$$

for $r=f$, $s=t$

$$P(R=r|s=t) = \alpha\Big(\big[0.5 \times 0.1 \times 0.2\,(0.9+0.1)\big] + \big[0.5 \times 0.5 \times 0.8\,(0.9+0.1)\big]\Big)$$

$$= \alpha\,(0.01 + 0.2) = \alpha\, 0.21$$

$$P(R|s=t) = \alpha \langle 0.09, 0.21 \rangle = \left\langle \frac{0.09}{0.3}, \frac{0.21}{0.3} \right\rangle$$

$$P(R|s=t) = \langle \underset{r=t}{0.3}, \underset{r=f}{0.7} \rangle$$

i.e. $P(Rain \mid Sprinkler = true) = \langle 0.3, 0.7 \rangle$

2. (2 points) Suppose your task is that of a 5-class classification of binary images of handwritten digits into categories $\{0, 1, 2, 3, 4\}$. The images have a resolution of 30×30 pixels, each that can take a value from $\{0, 1\}$. Say you are asked to design a Bayes classifier for this problem, but you are only allowed to use the raw image pixels as the input to the classifier (i.e., you cannot extract additional features from the images).

   (i) (0.5 points) Describe the classification rule that you will use.

   (ii) (0.5 points) What is the number of parameters that you will need for such a classifier?

   (iii) (0.5 points) Write the relationship between the Bayes classifier and the Naïve Bayes classifier.

   (iv) (0.5 points) If you design a Naïve Bayes classifier, what would be the number of parameters needed?

(i) Classification rule

$$\hat{y}_i = \underset{y \in \{0,1,2,3,4\}}{\text{Argmax}} \; P(y|x_i)$$

where $P(y|x_i) = P(x_i|y)\,P(y)$

(ii) To fully define the classifier, we need to define the distributions $P(y)$ and $P(x|y)$ for all values of y. For each y, the distribution $P(x|y)$ has $2^{(30\times30)} = 2^{900}$ possibilities, corresponding to each configuration of the $900 = 30\times30$ pixels & we need $2^{900}-1$ probabilities to fully describe the PMF.

Since there are 5 classes, we need a total of $5\times(2^{900}-1)$ parameters for $P(X|Y)$ and 4 params for $P(Y)$.

(iii) In Naïve Bayes, the features are assumed to be conditionally independent given the class label, i.e.,

$$P(X_1, X_2, X_3, \dots, X_n | Y) = \prod_{i=1}^{n} P(X_i|Y)$$

(iv) For each Y, we need $P(X_i|Y)\big|_{i=1,2,\dots n}$, i.e. n parameters for each Y. Since we have 5 classes, we need 5×900 for $P(X|Y)$ & 4 for $P(Y)$ i.e. 4504 parameters.

3. (2 points) Suppose we play n matches and each match is lost with a probability p independent of the results of other matches. Let K equal the number of losses in the n tests. Find the PMF $P_K(k)$.

See example 2.13 in Yates & Goodman for a soln. from first principles.
The PMF is that of a Binomial random variable

$$P_K(k) = \text{Binomial}(n, p)$$
$$= \binom{n}{k} p^k (1-p)^{n-k}$$

where $n \geq 1$ is an integer, $0 < p < 1$ and $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

and $\binom{n}{k} = 0$ for $k \notin \{0, 1, 2, \ldots, n\}$

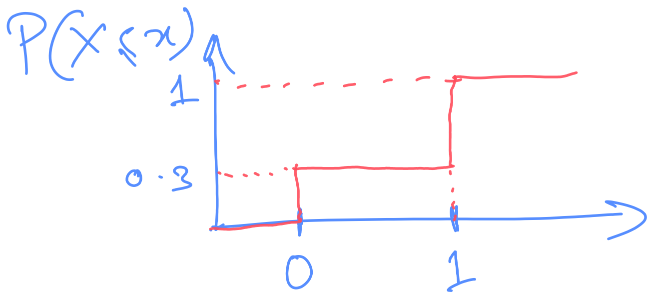
4. (3 points) [*Extra Credit*] Assume that you know how to generate samples from either a uniform distribution $\mathcal{U}(0,1)$ or a standard normal distribution $\mathcal{N}(0,1)$, write the algorithm you would follow to sample from

   (i) a distribution corresponding to a biased coin with a probability of heads being $p = 0.7$.

   (ii) a distribution corresponding to a Gaussian distribution with mean b and variance a^2.

(i) Use the CDF for a biased coin

Let 0 correspond to Tails & 1 correspond to Heads



Algorithm for sampling from a distribution corresponding to a biased coin:

1) Generate a sample z from $U(0,1)$

2) If $z \leq 0.3$
   output $x = 1$

   else
   output $x = 0$

The above algorithm will generate samples such that the $P_X(X=1) = 0.7$ & $P_X(X=0) = 0.3$ as required by the question.

The proof below is not expected

(ii) Algorithm

1) Let z be drawn from $N(0,1)$

2) output $x = az + b$

$$\mu = E[x] = E[az+b] = aE[z] + b = a \cdot 0 + b = b \Rightarrow E[x] = \mu_x = b$$

$$\sigma_x^2 = E[x^2] - (E[x])^2 = E[(az+b)^2] - (b)^2 = E[a^2z^2 + 2azb + b^2] - b^2$$

$$= E[a^2z^2 + 2abz] + \cancel{b^2} - \cancel{b^2} = a^2E[z^2] + 2abE[z] = a^2E[z^2] + 0$$

$$= a^2 \cdot 1$$

$$\{ \because E[z] = 0 \ \& \ E[z^2] = \sigma_z^2 + (E[z])^2 = \sigma_z^2 + 0 = \sigma_z^2 = 1$$