shashank chary Avusali

ARTIFICIAL INTELLIGENCE 2 – HOMEWORK 4

1. (4 points) Candy comes in two flavors: cherry and lime. Candy is wrapped and we can't tell which flavor until opened. There are 3 kinds of bags of candy:
   - H1= 75% cherry, 25% lime
   - H2= 50% cherry, 50% lime
   - H3= 25% cherry, 75% lime

   The hypothesis prior over H1, H2, and H3 is {0.4, 0.2, 0.4}. Given a new bag of candy,

   a. if we draw one candy and it is cherry. Using Bayesian learning, what's the probability that the next candy drawn from the bag is also cherry?

   Given that $d_1 = cherry$,

$$P(h_1|d_1 = cherry) = \alpha P(d_1 = cherry|h_1) \times P(h_1) = \alpha 0.75 \times 0.4 = 0.3\alpha$$

$$P(h_2|d_1 = cherry) = \alpha P(d_1 = cherry|h_2) \times P(h_2) = \alpha 0.5 \times 0.2 = 0.1\alpha$$

$$P(h_3|d_1 = cherry) = \alpha P(d_1 = cherry|h_3) \times P(h_3) = \alpha 0.25 \times 0.4 = 0.1\alpha$$

$$\alpha = \frac{1}{0.3 + 0.1 + 0.1} = 2$$

$$\therefore P(h_1|d_1 = cherry) = 0.6, P(h_2|d_1 = cherry) = 0.2, P(h_3|d_1 = cherry) = 0.2$$

$$P(d_2 = cherry|h_1, d_1 = cherry) = \sum_i P(d_2 = cherry|h_i) \times P(h_i|d_1 = cherry)$$

$$P(d_2 = cherry|h_1, d_1 = cherry) = 0.75 \times 0.6 + 0.5 \times 0.2 + 0.25 \times 0.2 = 0.6$$

   b. if we draw one candy and it is cherry. Using maximum a posteriori (MAP) learning, what's the probability that the next candy drawn from the bag is also cherry?

   Based on MAP, we would select the $H = h_1$.

$$P(d_2 = cherry) = 0.75$$

   c. if we draw two candies and they are both cherry. Using Bayesian learning, what's the probability that the next candy drawn from the bag is also cherry?

   Given that $d_1 = cherry, and\ d_2 = cherry$

$$P(h_1|d_2 = cherry) = \alpha P(d_2 = cherry|h_1) \times P(h_1|d_1 = cherry) = \alpha 0.75 \times 0.6 = 0.45\alpha$$

$$P(h_2|d_2 = cherry) = \alpha P(d_2 = cherry|h_2) \times P(h_2|d_1 = cherry) = \alpha 0.5 \times 0.2 = 0.1\alpha$$

$$P(h_3|d_2 = cherry) = \alpha P(d_2 = cherry|h_3) \times P(h_3|d_1 = cherry) = \alpha 0.25 \times 0.2 = 0.05\alpha$$

$$\alpha = \frac{1}{0.45 + 0.1 + 0.05} = 1.667$$

$$\therefore P(h_1|d_2 = cherry) = 0.75, P(h_2|d_2 = cherry) = 0.1667, P(h_3|d_1 = cherry) = 0.0833$$

$$P(d_3 = cherry|h_1, d_2 = cherry) = \sum_i P(d_3 = cherry|h_i) \times P(h_i|d_2 = cherry)$$

$$P(d_3 = cherry|h_1, d_2 = cherry) = 0.75 \times 0.75 + 0.5 \times 0.1667 + 0.25 \times 0.0833 = 0.6666$$

d. if we draw two candies and they are both cherry. Using maximum a posteriori (MAP) learning, what's the probability that the next candy drawn from the bag is also cherry?

Based on MAP, we would select the $H = h_1$.

$$P(d_3 = cherry) = 0.75$$

2. (1 point) Given the following data X, using the ML Gaussian estimator, what is $\boldsymbol{\mu}$? What is σ^2?

X: 1, 2, 3, 4, 6, 8

$$\mu = \frac{1}{n}\sum_{j=1}^{n} x_i \qquad \sigma^2 = \frac{1}{n}\sum_{j=1}^{n} (x_i - \mu)^2$$

$$\mu = \frac{1 + 2 + 3 + 4 + 6 + 8}{6} = 4$$

$$\sigma^2 = \frac{1}{6}(9 + 4 + 1 + 0 + 4 + 16) = \frac{34}{6} = 5.666$$

3. (3 points) Given the following Bayesian network model for the case of candies with an unknown proportion of cherries and limes, where the wrapper color (red or green) probabilistically depends on the candy flavor. Suppose we unwrap 10 candies and they are as follows.

(cherry, red), (cherry, green), (lime, red), (lime, red), (cherry, red), (lime, green), (cherry, green), (cherry, red), (lime, red), (cherry, red). What are the maximum likelihood estimation (MLE) of parameters θ, $\theta1$, and $\theta2$?

$$\theta = \frac{c}{c + l} = \frac{6}{10} = 0.6$$

$$\theta_1 = \frac{r_c}{r_c + g_c} = \frac{4}{6} = \frac{2}{3}$$

$$\theta_2 = \frac{r_l}{r_l + g_l} = \frac{3}{4}$$

4. Find and read a paper that applies HMM and particle filter algorithm to solve some problems and write a 1-page review. Specifically, the requirements are as follows:

A Comparison Study of Hidden Markov Model and Particle Filtering Method: Application to Fault Diagnosis for Gearbox

IEEE Transactions on Prognostics & System Health Management Conference 2012

Yunxian Jia, Lei Sun, Hongzhi Teng

The paper compares two data driven models Hidden Markov Model (HMM), and Particle Filter(PF) for fault diagnosis and fault identification in gearbox. The experiments demonstrate that the PF has better accuracy over HMM, but HMM was computationally efficient among the two. Part of the reason could be the inability of HMM to handle continuous state space, and ability of PF to handle discrete and continuous state spaces.

In general, a gearbox lifecycle from normal to failure follows a gradual degradation multi state process, which can be judged by indirect parameters such as vibration signals, which can be measured through sensors. Multiple HMMs are trained to recognize the states of a machine, one HMM per state. A HMM is trained for each distinct state of interest with labelled set of sequences of that state. For testing an observed sequence is fed into each HMM, current machine state is identified based on the HMM that produces maximum probability. They could have extracted the features based on the observed sequence and fed into a classifier to detect the state of the machine, but the authors claim that the features would represent a limited characteristics of an observed sequence, and would degrade the accuracy of the model.

The authors did not do a good job at specifying the details of HMM, PF models that are used in the paper. Their results on simulated setup demonstrated that, HMM achieved 92% accuracy where as PF achieved an accuracy of 98%. For a realtime test bed setup, they ran a mechanical motor for 450 hours, vibration signals are collected from the motor every 5 minutes. The authors have trained 5 HMMs, their transition probabilities are calculated using forward and backward algorithm.

Overall, the paper was not an easy read. More details on number of evidences per hidden node, and details about the particle filter parameters that they used would have been much appreciated.