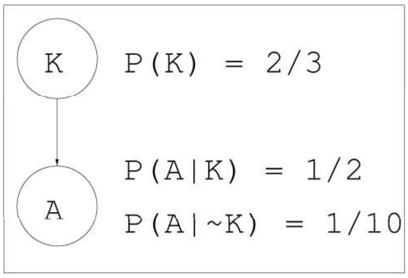
Subject Name : ML, End Sem Examination, 40 Points, 29/10/2023

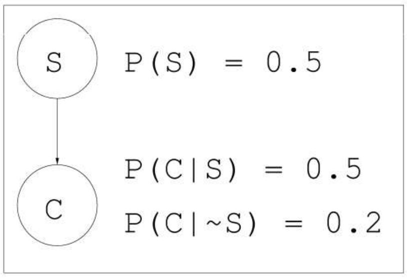
Q1 [10 POINTS] : BAYESIAN INFERENCE

(a) **Kangaroos.**



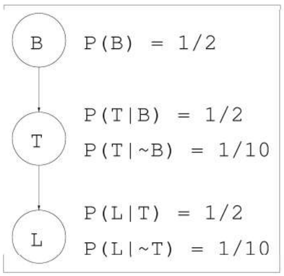
Half of all kangaroos in the zoo are angry, and 2/3 of the zoo is comprised of kangaroos. Only 1 in 10 of the other animals are angry. What's the probability that a randomly-chosen animal is an angry kangaroo?

(b) **Stupidity.**



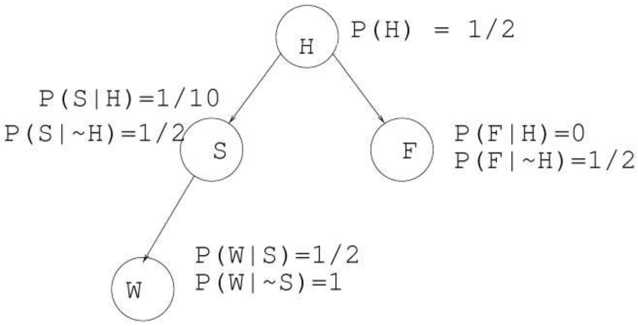
Half of all people are stupid. If you're stupid then you're more likely to be confused. A randomly-chosen person is confused. What's the chance they're stupid?

(c) **Potatoes.**



Half of all potatoes are big. A big potato is more likely to be tall. A tall potato is more likely to be lovable. What's the probability that a big lovable potato is tall?

(d) **Final part.**



What's $P(W \wedge F)$?

Q2 [10 Points] : **Generative Classifier :**

Let us consider the IRIS dataset where we focus on X variable SepalLength and our target variable Species. In this dataset, there are three classes Y : Setosa, Versicolor, and Virginia which have equiprobability as their priors. Fit the Gaussian Distribution for each class using (5, 0.34) , (5.93, 0.51) and (6.98,0.62) as mean and variance respectively for Setosa, Versicolor and Virginia. Predict the posterior probability P(X=7, Y=SETOSA), P(X=7, Y=VERSICOLOR), and P(X=7, Y=VIRGINIA). Please write the Python code to perform the same and verify your results as computed aforementioned. Please write full steps in justification of your code and verification of your results.

Q3 [5 Points] **Regression** :

We are dealing with samples x where x is a single value. We would like to test two alternative regression models:

1. $y = ax + e$
2. $y = ax + bx^2 + e$

We make the same assumptions we had in class about the distribution of e ($e \sim N(0,s^2)$).

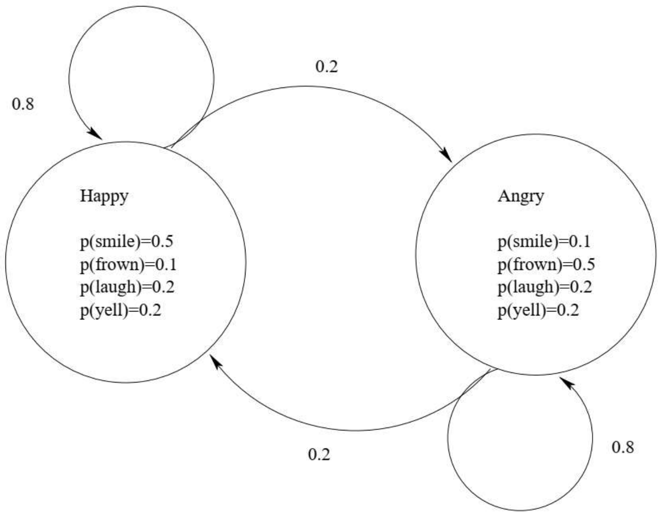
a. (4 points) Assume we have n samples: $x_1 \ldots x_n$. with their corresponding y values, : $y_1 \ldots y_n$. Derive the value assigned to b in model 2. You can use a in the equation for b.

b. (2 points) Which of the two models is more likely to fit the *training* data better?

   a. model 1
   b. model 2
   c. both will fit equally well
   d. impossible to tell

Q4 [10 Points] : HMM

Andrew lives a simple life. Some days he's Angry and some days he's Happy. But he hides his emotional state, and so all you can observe is whether he smiles, frowns, laughs, or yells. We start on day 1 in the Happy state, and there's one transition per day.



Definitions:

q_t = state on day t.

O_t = observation on day t.

(a) What is $P(q_2 = Happy)$?

(b) What is $P(O_2 = frown)$?

(c) What is $P(q_2 = Happy | O_2 = frown)$?

(d) What is $P(O_{100} = yell)$?

(e) Assume that $O_1 = frown$, $O_2 = frown$, $O_3 = frown$, $O_4 = frown$, and $O_5 = frown$. What is the most likely sequence of states?

Q5 : [5 POINTS] Prove that in Bayesian Logistic Classification where cross entropy is used as a loss function, the gradient of error function (i.e negative of log likelihood function P(t/w)) is equal to product of basis function (ϕ) and error between original samples “t” and its prediction “y”.