6

Quiz 5: Bayes Rule

Student Name:

Student Number:

Suppose we are modelling real-valued data as coming from a Gaussian distribution with unit variance and mean given by a parameter μ. Suppose we use a prior distribution on the parameter μ that is Gaussian with mean zero and unit variance. We observe N data points $x_1, x_2, x_3, \dots x_N$.

$lh = N(\mu, 1)$

$P(\mu) = N(0, 1)$.

You can use $N(x; \mu, s^2)$ as the pdf of the Gaussian distribution.

a) Write an expression for the posterior distribution over μ, in terms of $x_1, x_2, x_3, \dots x_N$.

$$P(\mu \mid x_1 x_2 \cdots x_N) = \frac{P(\mu, x_1 x_2 \cdots x_N)}{p(x_1 x_2 \cdots x_N)}$$

$$\propto P(x_1 x_2 \cdots x_N \mid \mu)\, P(\mu)$$

$$\propto \left[\prod_{i=1}^{N} N(x_i; \mu, 1) \right] N(0,1)$$

$$\propto \left\{ \prod_{i=1}^{N} \exp\left[-\frac{(x_i - \mu)^2}{2} \right] \right\} \exp\left[-\frac{\mu^2}{2} \right]$$

$$= \exp\left[-\frac{\prod_{i=1}^{N} (x_i - \mu)^2}{2} - \frac{\mu^2}{2} \right]$$

$$\propto \exp\left[-\frac{1}{2}\mu^2 (N+1) + \mu \left(\prod_{i=1}^{N} x_i \right) \right].$$

we know if $y \sim N(a, b)$

then $p(y) \propto \exp\left[-\frac{(y-a)^2}{2b} \right] \propto \exp\left[-y^2 \left(\frac{1}{2b}\right) + y\left(\frac{a}{b}\right) \right]$

$$\begin{cases} \frac{N+1}{2} = \frac{1}{2b} \\ \prod_{i=1}^{N} x_i = \frac{a}{b} \end{cases} \Rightarrow \begin{cases} a = \frac{1}{N+1} \prod_{i=1}^{N} x_i \\ b = \frac{1}{N+1} \end{cases}$$

so $P(\mu \mid x_1 x_2 \cdots x_N) \sim N\left(\frac{1}{N+1} \prod_{i=1}^{N} x_i, \frac{1}{N+1} \right)$.

b) Is the prior conjugate?

Yes, because the posterior is of the same form as the prior.

10.

Quiz 6: Bayesian Inference

Student Name:

Student Number:

We are measuring the temperature of a bottle of water. The temperature is denoted as θ. At one atmospheric pressure, temperature of water ranges between 0 and 100 Celsius. Our prior guess for the water temperature is a uniform distribution. Our noisy measurement, x, is probabilistic based on θ. The pdf is a triangle centered at θ with a width of 2W. $p(x|\theta)=\frac{1}{W}\left(1-\frac{1}{W}|x-\theta|\right)$ for $|x-\theta|<W$ and $p(x|\theta)=0$ otherwise. Measurements are independent conditioned on θ.

a. Under hypothesis 1 (H1), W=10. What is the range of values that the first observation x_1 may take on?

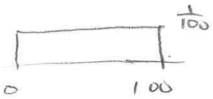
$\theta \in [0, 100]$

$x_1 \in [\theta - w, \theta + w]$

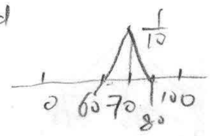
$x_1 \in [0-10, 100+10] = [-10, 110]$

b. We made the first observation: $x_1=70$. Under H1, what is the range of values that the second observation may take on?

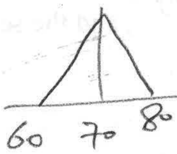
prior



likelihood



posterior (θ').



$\theta' \in [60, 80]$

$x_2 \in [\theta' - w, \theta' + w]$

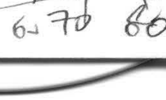
$\in [60-10, 80+10] = [50, 90]$

$h = \frac{1}{10}$

c. We made the first observation: $x_1=70$. Under H1, what is the expected value of θ?

$E[\theta | x_1 = 70] = \int_\theta P(\theta | x_1 = 70) \, d\theta$

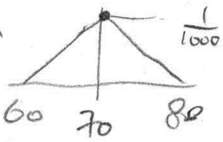
$= \int_\theta \underbrace{P(\theta) P(x_1 = 70|\theta) / P(x_1=70)}_{\text{posterior}} \, d\theta = 70$



Symmetric

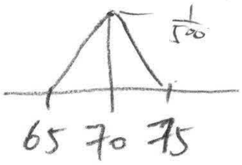
d. Under hypothesis 2 (H2), W=5. H1 and H2 are equally likely *a prior*. After observing the first data point $x_1 = 70$, what is posterior probability of H1 and H2?

$$P(x_1=70|H_1) = \int_{\theta_1} P(x_1=70|\theta_1)P(\theta_1)\,d\theta_1$$

$= \text{area of}$ 

$= \frac{1}{2} \times 20 \times \frac{1}{1000} = \frac{1}{100}$

$$P(x_1=70|H_2) = \int_{\theta_2} P(x_1=70|\theta_2)P(\theta_2)\,d\theta_2$$

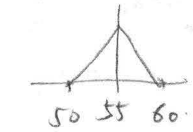
$= \text{area of}$ 

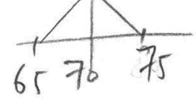
$= \frac{1}{2} \times 10 \times \frac{1}{500} = \frac{1}{100}$

also $P(H_1) = P(H_2) = \frac{1}{2}$

So $P(H_1|x_1=70) = \frac{1}{2}$ $P(H_2|x_1=70) = \frac{1}{2}$

e. Again, H1 and H2 are equally likely *a prior*. After observing the first data point $x_1 = 70$, and the second datapoint $x_2 = 55$, what is posterior probability of H1 and H2?

for $x_2 = 55$, $P(x_2=55|H_2) =$ 

but $P(x_1=70|H_2) =$ 

Since two observation are independent, their likelihood multiplies. this implies that $P(x_1=70, x_2=55|H_2) = 0$.

So $P(H_1|x_1=70, x_2=55) = 1$

and $P(H_2|x_1=70, x_2=55) = 0$.

Quiz 7: Naïve Bayes

Student Name:

Student Number:

We are predicting if it will snow in the afternoon by whether it snowed in the morning. x=0 indicates that it did not snow in the morning while x=1 indicates that it did. Similarly, y indicates whether it snows in the afternoon. Both of these variables are binary. We collected a dataset of 100 days (data points). In this dataset, x=y=1 for 30 days, x=y=0 for 40 days, x=1,y=0 for 20 days, and x=0, y=1 for 10 days.

(1) Suppose we apply a naïve Bayes model to this dataset. Compute the probability that it will snow in the afternoon given if it snowed in the morning (please give numerical values for P(y=1|x=0) and P(y=1|x=1)).

	x=0	x=1
y=0	40	20
y=1	10	30
total	50	50

$p(y=1|x=0) = \frac{10}{50} = \frac{1}{5}$

$p(y=1|x=1) = \frac{30}{50} = \frac{3}{5}$

Only one binary feature, no assumption is made by Naive Bayes.

(2) This naïve Bayes model is the same as a logistic regression model with a bias and the feature x. Note that the prediction of logistic regression is: $p(y = 1) = \sigma(\theta_0 + \theta_1 x)$. What is the numerical value of θ_0 and θ_1 such that the logistic regression model is the same as the naïve Bayes model?

$$\sigma(\theta_0+\theta_1 x) = p(y=1)$$

for $x=0$,

$$\sigma(\theta_0+\theta_1 \cdot 0) = \frac{1}{5}$$

$$\Rightarrow \frac{1}{1+e^{-\theta_0}} = \frac{1}{5} \qquad 5 = 1+e^{-\theta_0} \qquad e^{-\theta_0} = 4 \qquad \theta_0 = -\log 4.$$

for $x=1$,

$$\sigma(\theta_0+\theta_1 \cdot 1) = \frac{3}{5}$$

$$\Rightarrow \frac{1}{1+e^{-\log 4-\theta_1}} = \frac{3}{5}$$

$$3+3e^{-\log 4-\theta_1} = 5$$

$$e^{-\log 4-\theta_1} = \frac{2}{3}$$

$$\log 4+\theta_1 = \log\frac{3}{2}$$

$$\theta_1 = \log\frac{3}{2} - \log 4 = \log\frac{3}{8}$$

6

Quiz 8: Bayesian Networks

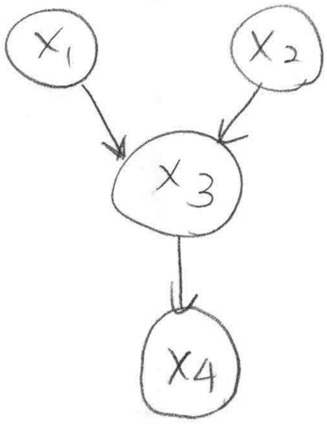
Student Name:

Student Number:

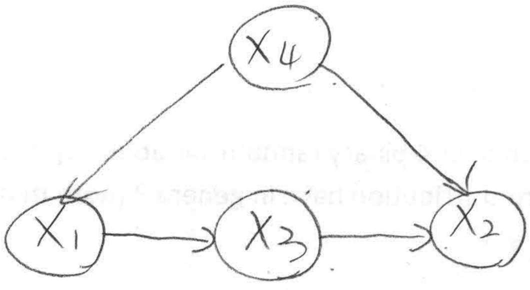
Consider a general distribution over 4 binary random variables: x_1, x_2, x_3, x_4. How many free parameters does the joint distribution have in general? (note that a proper distribution is normalized)

$2^4 - 1$

Suppose we know that x_1 is independent of x_2 when no other variables are observed. Draw a Bayesian network that has this property.



Draw another Bayesian network where x_1 and x_2 are dependent when no other variables are observed, ~~when x_3 is observed and when x_4 is observed,~~ but where they become **independent** when **both** x_3 and x_4 are observed.



6

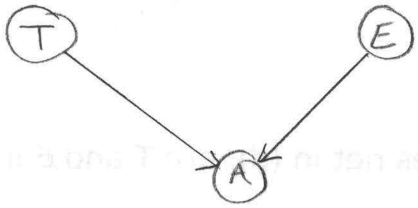
Quiz 9: Conditional independence

Student Name:

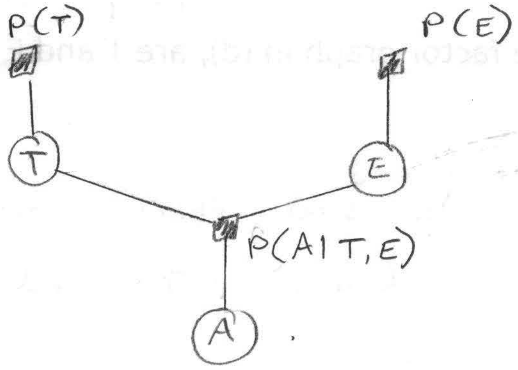
Student Number:

We consider three binary variables: truck hit house (T), earthquake (E) and house alarm goes off (A). It is assumed that T and E are independent without observing A. A depends on both T and E.

(a) Draw the Bayesian network.



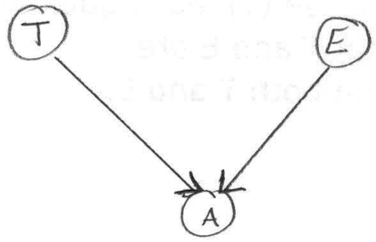
(b) Draw an (undirected) factor graph and write down an expression for each local function using with three local functions P(T), P(E) and P(A|T,E).

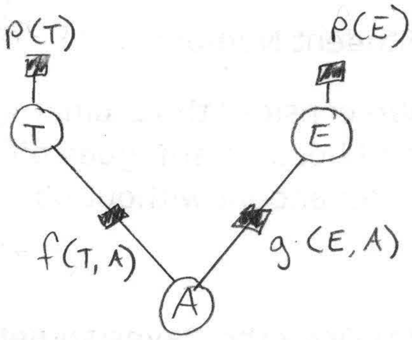


(c) From the factor graph you drew in part (b), can you conclude that T and E are independent without observing A? (Your answer should be based only on the graph, not on the functions themselves.)

No, since there exists a direct path between T and E.

(d) Suppose we know that P(A|T,E)=f(T,A)g(E,A). Draw a Bayesian network and a factor graph for the resulting model.





(e) Based on the Bayes net in (d), are T and E independent after observing A?

No, since observing A allows us to travel along the edges from T to E.

(f) Based on the factor graph in (d), are T and E independent after observing A?

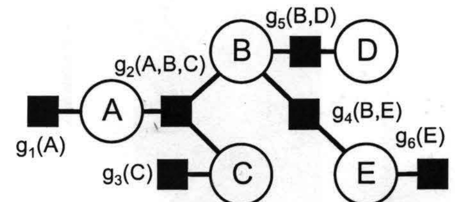
Yes, since there exists no path between T and E.

ECE521: Quiz 10

Student Name:

Student Number:

Consider the following factor graph, which describes an *unnormalized* distribution over binary variables:



A	$g_1(A)$	C	$g_3(C)$	E	$g_6(E)$	BE	$g_4(B,E)$	BD	$g_5(B,D)$	ABC	$g_2(A,B,C)$
0	1	0	3	0	2	00	1 × 2	00	3 646	000	3 × 3
1	3	1	2	1	5	01	3 × 5	01	2 646	001	4 × 2
						10	2 × 2	10	1 1632	010	2 × 3
						11	4 × 5	11	5 1632	011	4 × 2
										100	1 × 9
										101	2 × 6
										110	4 × 9
										111	3 × 6

> 17 > 14 > 21 > 54 38 68

1. Use the sum-product algorithm to compute the following messages:

$\mu_{g_1 \to A}(0) =$ 1 $\qquad \mu_{g_1 \to A}(1) =$ 3

$\mu_{A \to g_2}(0) =$ 1 $\qquad \mu_{A \to g_2}(1) =$ 3

$\mu_{g_3 \to C}(0) =$ 3 $\qquad \mu_{g_3 \to C}(1) =$ 2

$\mu_{C \to g_2}(0) =$ 3 $\qquad \mu_{C \to g_2}(1) =$ 2

(B) $\mu_{g_2 \to B}(0) =$ 38 $\qquad \mu_{g_2 \to B}(1) =$ 68

$\mu_{g_6 \to E}(0) =$ 2 $\qquad \mu_{g_6 \to E}(1) =$ 5

$\mu_{E \to g_4}(0) =$ 2 $\qquad \mu_{E \to g_4}(1) =$ 5

(B) $\mu_{g_4 \to B}(0) =$ 17 $\qquad \mu_{g_4 \to B}(1) =$ 24

$\mu_{B \to g_5}(0) = 38 \times 17 = 646$ $\qquad \mu_{B \to g_5}(1) = 68 \times 24 = 1632$

$\mu_{g_5 \to D}(0) = 3 \times 646 + 1 \times 1632$ $\qquad \mu_{g_5 \to D}(1) = 2 \times 646 + 5 \times 1632$

$= 3570$ $\qquad = 9452$

2. Based on the message arriving at D, what is the normalizing constant of the factor graph? That is, what constant value should the product of local functions be multiplied by, so that the distribution is normalized?

$$Z = \mu_{g_5 \to D}(0) + \mu_{g_5 \to D}(1)$$
$$= 13022.$$

3. Based on the message arriving at D, what is the marginal probability $P(D)$ for $D = 0$ and $D = 1$?

$$P(D=0) = \frac{\mu_{g_5 \to D}(0)}{Z} = \frac{3570}{13022} = 0.27$$

$$P(D=1) = \frac{\mu_{g_5 \to D}(1)}{Z} = \frac{9452}{13022} = 0.73.$$

4. Suppose we observe that $E = 1$. Below, modify the message updates that you derived above so as to compute the posterior marginal probability $P(D|E = 1)$ for $D = 0$ and $D = 1$.

$$\mu_{E \to g_4}(0,1) = [0 \ 5]$$

$$\mu_{g_4 \to B}(0) = 1 \times 0 + 3 \times 5 = 15 \qquad \mu_{g_4 \to B}(1) = 2 \times 0 + 4 \times 5 = 20.$$

same as before

$$\mu_{B \to g_5}(0,1) = \mu_{g_4 \to B} \cdot \mu_{g_2 \to B} = [38 \times 15, \ 68 \times 20] = [570 \ 1360]$$

$$\mu_{g_5 \to D}(0) = 3 \times 570 + 1 \times 1360 = 3070$$

$$\mu_{g_5 \to D}(1) = 2 \times 570 + 5 \times 1360 = 7940.$$

$$P(D=0|E=1) = \frac{3040}{3040+7940} = 0.28 \qquad P(D=1|E=1) = \frac{7940}{3040+7940} = 0.72$$