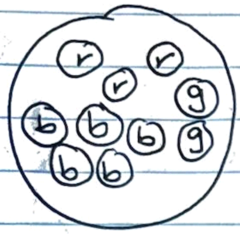
Courtney Hodge

Bayesian ML - Homework 1 - Probability and Priors

Problem 1: Basic Probability



10 balls total — 5/10 blue
3/10 red
2/10 green

(a) The probability of drawing a red ball from the bag is 3/10

~~5/10 blue balls~~
(b) ~~4/9 blue balls~~ ~~Probability that~~

5/10 probability that the first ball is blue
4/9 probability that the next ball is also blue

$\frac{5}{10} \times \frac{4}{9} = \frac{20}{90} = \frac{2}{9}$, probability that the next ball is also blue

Problem 2: Independent Events

P(Server being down) = 0.05, server failures are independ.

Probability of 2 independent events — P(A and B) = P(A) x P(B)

P(server 1 being down AND server 2 being down) =
P(0.05) x P(0.05) = 0.0025

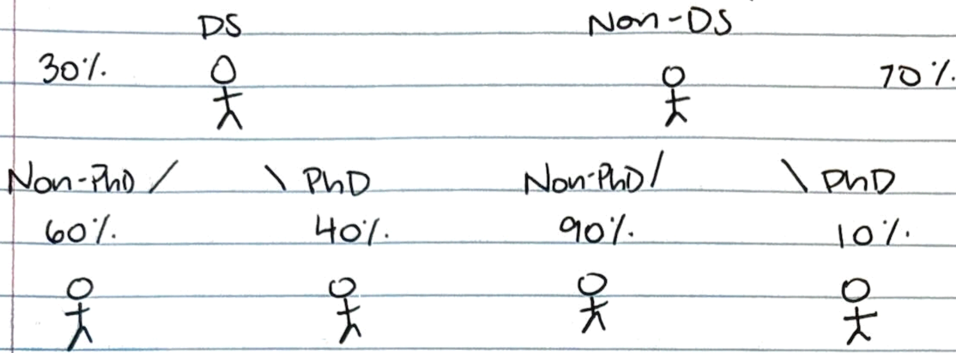
(a) 0.0025

Probability of at least 1 being down
P(server 1 being down OR server 2 being down) =
P(0.05) + P(0.05) = 0.1

(b) 0.1

Courtney Hodge

Problem 3: Conditional Probability



(a) If an employee is chosen randomly, what is the probability that the employee is a Data Scientist?

$$P(DS) \times P(PhD \mid DS) = \left(\frac{3}{10}\right) \times \left(\frac{4}{10}\right) = \frac{12}{100}$$

(b) Given that an employee has a PhD, what is the probability that the employee is a Data Scientist?

Bayes' Theorem

$$P(A \mid B) = \frac{P(B \mid A) P(A)}{P(B)} \rightarrow P(DS \mid PhD) = \frac{P(PhD \mid DS) P(DS)}{P(PhD)}$$

#1 Calculate PhD using the law of total probability:

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i) P(B_i)$$

$$P(PhD) = P(PhD \mid DS) P(DS) + P(PhD \mid Non\text{-}DS) P(Non\text{-}DS)$$

$$= \left(\frac{4}{10}\right)\left(\frac{3}{10}\right) + \left(\frac{1}{10}\right)\left(\frac{7}{10}\right)$$

$$= \frac{12}{100} + \frac{7}{100} = \frac{19}{100}$$

#2 Solve: $P(DS \mid PhD) = \dfrac{\left(\frac{4}{10}\right)\left(\frac{3}{10}\right)}{\left(\frac{19}{100}\right)} = 0.63$

Cortney Hodge

Problem 4: Law of Total Probability

Law of Total Probability

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

P(Tested Positive | Has Disease) = 0.95

P(Tested Positive | Doesn't Have Disease) = 0.1

P(Has Disease) = 0.5% = 0.005

P(Doesn't Have Disease) = 0.995

(a) If a person tested positive in the test, what is the probability that the person actually has the disease?

$$\underbrace{P(\text{Has Disease} \mid \text{Tested Positive})}_{P(A \mid B)} = \frac{\underbrace{P(\text{Has Disease})}_{P(A)}\ \underbrace{P(\text{Tested Positive} \mid \text{Has Disease})}_{P(B|A)}}{\underbrace{P(\text{Tested Positive})}_{P(B)}}$$

$$P(A|B) = \frac{0.005(0.95)}{P(B)} = \frac{0.005(0.95)}{0.10425} = 0.0456$$

$P(B) = P(\text{Tested Positive}) = P(\text{Tested Pos} \mid \text{Has Disease})\,P(\text{Has Disease}) + P(\text{Tested Pos} \mid \text{Doesn't Have Disease}) * P(\text{Doesn't Have Disease})$

$$P(\text{Tested Positive}) = 0.95(0.005) + (0.1)0.995 = 0.10425$$

(b) What is the total probability of a person testing positive?

$$P(\text{Tested Positive}) = 0.10425$$

Courtney Hodge

Problem 5

Bayes' Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

P(Spam) = .9

P(Not-Spam) = 0.1

P(Identified Not-Spam | Spam) = 0.05

P(Identifies Spam | Spam) = 0.95

P(Identifies Not-Spam | Not-Spam) = 0.85

P(Identified Spam | Not-Spam) = 0.15

(a) If an email is picked at random, and the filter classifies it as spam, what is the probability that it is actually spam?

~~P(Spam) x P(Identifies Spam | Spam) = 0.9 x 0.95 = 0.855~~

P(B | A) P(A)

$$\underbrace{P(\text{Spam}}_{P(A} \mid \underbrace{\text{Identified Spam})}_{B)} = \frac{P(\text{Identified Spam}|\text{Spam})P(\text{Spam})}{\underbrace{P(\text{Identified Spam})}_{P(B)}}$$

$$P(A|B) = \frac{0.95(0.9)}{P(B)} = \frac{0.95(0.9)}{0.87} = 0.983$$

Law of Total Probability

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

P(Identified Spam) = P(Identified Spam | Spam)P(Spam) + P(Identified Spam | Not-Spam)P(Not-Spam)

= 0.95(.9) + (0.15) 0.1

= 0.87

P(Identified Spam | Not-Spam) = 0.15

Courtney Hodge

Problem 5

(b) If an email is classified as "not spam", what is the probability that it is actually spam?

P(~~am~~ Spam | ~~Cla~~ Identified Not-Spam)

$$P(A|B) = \frac{P(A)\,P(B|A)}{P(B)}$$

$$= \frac{P(\text{Spam})\,P(\text{Identified Not-Spam} \mid \text{Spam})}{P(\text{Identified Not-Spam})}$$

$$= \frac{0.9\,(0.05)}{P(\text{Identified Not-Spam})} = \frac{0.9(0.05)}{0.13} = 0.346$$

Law of Total Probability

$$P(A) = \sum_{i=1}^{n} P(A|B_i)\,P(B_i)$$

$$P(\text{Identified NOT Spam}) = P(\text{Identified NOT Spam} \mid \text{Spam})\,P(\text{Spam}) + P(\text{Identified Not Spam} \mid \text{Not-Spam}) * P(\text{Not-Spam})$$

$$= 0.05\,(0.9) + (0.85)\,0.1$$

$$= 0.13$$

Courtney Hodge

Problem 6: Expectation of a Discrete Random Variable

(a) Define the random vairable x that models this game.

$\frac{1}{6} = 6$, win \$10

$\frac{5}{6}$ = 1, 2, 3, 4 or 5 , lose \$2

$\cancel{P(\frac{1}{6}) + P(\frac{5}{6})}$

X = 10 , if you roll a 6
X = -2 , if you roll 1 through 5

$$X = \begin{cases} 10, & \text{if a 6 appears} \\ -2, & \text{if any number 1 to 5 appears} \end{cases}$$

(b) Compute the expected value of X

P(if a 6 appears)X + P(if you roll 1 → 5)X

To compute the expected value of X, we use the expected value of a discrete random variable

$E(x) = \sum_i P(X_i) \cdot X_i$, where X_i are the possible outcomes and $P(X_i)$ are the corresponding probabilities.

Prob of rolling a 6 is $P(X=10) = 1/6$
Prob of rolling anything else $P(X=-2) = 5/6$

$E(X) = P(X=10)(10) + P(X=-2)(-2)$ $E(X) = 0$

$\frac{-10}{6} = 0$

$(\frac{1}{6})(10) + (5/6)(-2) = \frac{10}{6} + (-\frac{10}{6}) = 0$

Courtney Hodge

Problem 7: Expectation of a Continuous Random Variable

(a) Compute the expected value E[X] of X

Probability Density Function

$f(x) = 2e^{-2x}$ for $x \geq 0$

PDF

$E(x) = \int_{-\infty}^{\infty} x f(x)\,dx$

$E(x) = \int_{0}^{\infty} x(2e^{-2x})\,dx$

Integration By Parts

$u = x$ $du = dx$

$v = -e^{-2x}$ $dv = 2e^{-2x}$

$\int 2e^{-2x}\,dx$ $u = -2x$

$du = -2dx$

$-\int e^{u}\,du$ (exponential rule)

$= -e^{u}$ $\boxed{= -e^{-2x}}$

$\int u\,dv = uv - \int v\,du$ (take neg out)

$E(x) = \left[-xe^{-2x}\right]_0^{\infty} + \int_0^{\infty} e^{-2x}\,dx$

Solving, we get... At $x=0$, $-xe^{-2x} = 0$

At $x = \infty$, $-xe^{-2x}$ ~~goes to~~ $= 0$

So $\left[-xe^{-2x}\right]_0^{\infty}$ is $= 0$

Solving $\int_0^{\infty} e^{-2x}\,dx = \left[\frac{-1}{2}e^{-2x}\right]_0^{\infty} = \left(0 - \left(\frac{-1}{2}\right)\right) = \frac{1}{2}$

Therefore, $\boxed{E(X) = 0 + \frac{1}{2} = \frac{1}{2}}$

(b) Compute the variance Var[X] of X

Computing $E(X^2)$: $E(X^2) = \int_0^{\infty} x^2 \cdot f(x)\,dx$

Variance of a continuous random variable X is

$Var(X) = E(X^2) - (E(X))^2$

$E(x^2) = \int_0^{\infty} x^2(2e^{-2x})\,dx$

Since we know that the PDF of an exponential distribution is $f(x) = \lambda e^{-\lambda x}$, then... $E(X) = 1/\lambda$

$E(X^2) = 2/\lambda^2$ ... with $\lambda = 2$

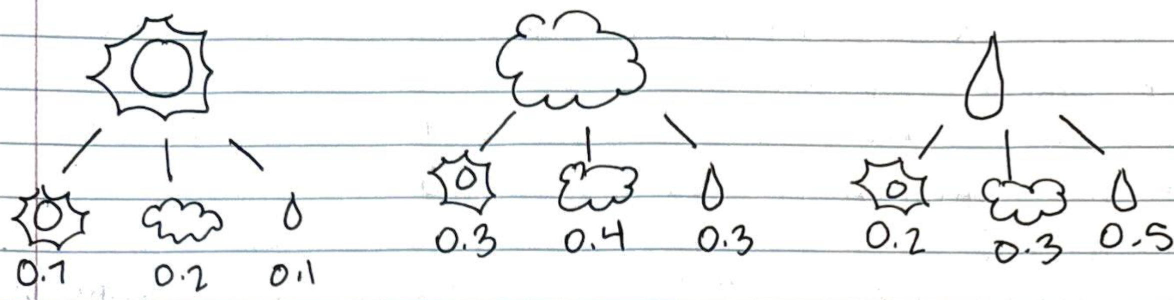
$E(X^2) = \frac{2}{\lambda^2} = \frac{2}{2^2} = \frac{2}{4} = \frac{1}{2}$ → Now we can compute the Variance

$V(X) = E(X^2) - (E(X))^2 = 0.5 - (0.5)^2 = 0.5 - 0.25 = 0.25$

$V(X) = 0.25$ hrs^2

Courtney Hoang

Problem 8: Markov Chain



(a) Construct the transition matrix of this Markov chain

$$P_r = \begin{array}{c} \\ \text{Sunny} \\ \text{Cloudy} \\ \text{rainy} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Sunny} & \text{cloudy} & \text{rainy} \end{array} \\ \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{bmatrix} \end{array}$$

(b) If today is sunny, what is the probability that it will be rainy two days from now?

*Compute the two-step transition matrix P^2

$P^2 = P \times P$

$P^2_{(1,1)} \rightarrow 0.7\times0.7 + 0.2\times0.3 + 0.1\times0.2 = 0.49 + 0.06 + 0.02 = 0.57$

$P^2_{(1,2)} \rightarrow 0.7\times0.2 + 0.2\times0.4 + 0.1\times0.3 = 0.14 + 0.08 + 0.03 = 0.25$

$P^2_{(1,3)} \rightarrow 0.7\times0.1 + 0.2\times0.$[illegible]$5 + 0.1\times0.5 = 0.07 + 0.06 + 0.05 = 0.18$

$P^2_{(2,1)} \rightarrow 0.3\times0.7 + 0.4\times0.3 + 0.3\times0.2 = 0.21 + 0.12 + 0.06 = 0.39$

$P^2_{(2,2)} \rightarrow 0.3\times0.2 + 0.4\times0.4 + 0.3\times0.3 = 0.06 + 0.16 + 0.09 = 0.31$

$P^2_{(2,3)} \rightarrow 0.3\times0.1 + 0.4\times0.3 + 0.3\times0.5 = 0.03 + 0.12 + 0.15 = 0.30$

$P^2_{(3,1)} \rightarrow 0.2\times0.7 + 0.3\times0.3 + 0.5\times0.2 = 0.14 + 0.09 + 0.1 = 0.33$

$P^2_{(3,2)} \rightarrow 0.2\times0.2 + 0.3\times0.4 + 0.5\times0.3 = 0.04 + 0.12 + 0.15 = 0.31$

$P^2_{(3,3)} \rightarrow 0.2\times0.1 + 0.3\times0.3 + 0.5\times0.5 = 0.02 + 0.09 + 0.25 = 0.36$

$$P^2 = \begin{array}{c} \\ S \\ C \\ R \end{array} \begin{array}{c} \begin{array}{ccc} S & C & R \end{array} \\ \begin{bmatrix} 0.57 & 0.25 & 0.18 \\ 0.39 & 0.31 & 0.30 \\ 0.33 & 0.31 & 0.36 \end{bmatrix} \end{array}$$

If today is sunny, the probability that it will be rainy two days from now is 0.18

Courtney Hodge

Problem 8: Markov Chain

(c) Find the stationary distribution of this Markov Chain

$$P = \begin{array}{c} \\ S \\ C \\ R \end{array} \begin{array}{c} \begin{array}{ccc} S & C & R \end{array} \\ \begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{pmatrix} \end{array}$$

* The stationary distr. ~~matrix~~ of a Markov Chain is a probability vector $\pi = [\pi_1, \pi_2, \pi_3]$ that remains unchanged after multiplying with the transition matrix P. $\rightarrow \pi P = \pi$. Also, since π is a probability vector, we know: $\pi_1 + \pi_2 + \pi_3 = 1$

This is equivalent to

$\pi_1 \times 0.7 + \pi_2 \times 0.3 + \pi_3 \times 0.2 = \pi_1$

$\pi_1 \times 0.2 + \pi_2 \times 0.4 + \pi_3 \times 0.3 = \pi_2$

$\pi_1 \times 0.1 + \pi_2 \times 0.3 + \pi_3 \times 0.5 = \pi_3$

$0.7\pi_1 + 0.3\pi_2 + 0.2\pi_3 = \pi_1$ Solve for 0

$-0.3\pi_1 + 0.3\pi_2 + 0.2\pi_3 = 0$

$0.2\pi_1 + 0.4\pi_2 + 0.3\pi_3 = \pi_2$ solve for 0

$0.2\pi_1 - 0.6\pi_2 + 0.3\pi_3 = 0$

$0.1\pi_1 + 0.3\pi_2 + 0.5\pi_3 = \pi_3$ solve for 0

$0.1\pi_1 + 0.3\pi_2 - 0.5\pi_3 = 0$

I followed these steps from chat GPT!

Solving each equation for π_1, π_2, and π_3, I got the stationary distribution as follows

$\pi_1 = 0.4545 \quad \pi_2 = 0.3182 \quad \pi_3 = 0.2273$

These values are the probabilities of being in each state (sunny, cloudy, rainy) after a long period of time, regardless of the starting weather.

Courtney Hodge

Problem 8: Markov Chain

(d) Interpret the stationary distribution in the context of this weather model

$\pi_1 = 0.4545$

$\pi_2 = 0.3182$

$\pi_3 = 0.2273$

$\pi_1 = 0.4545$: Over a long period of time, about 45.45% of the days will be sunny

$\pi_2 = 0.3182$: Over a long period of time, about 31.82% of the days will be cloudy

$\pi_3 = 0.2273$: Over a long period of time, about 22.73% of the days will be rainy

This is because the probabilities stabalize over time

Courtney Hodge

Problem 9: Conjugate Priors and Posterior Distributions

30 patients recover out of 100 patients

Bernoulli Distr.

$P(X|\theta) = \theta^{x} \cdot (1-\theta)^{1-x}$, where $x \in \{0,1\}$ and $0 \leq \theta$

Beta Distr.

$P(\theta|\alpha,\beta) = \dfrac{\theta^{(\alpha-1)} \cdot (1-\theta)^{(\beta-1)}}{B(\alpha,\beta)}$, where $B(\alpha,\beta)$ is the beta function, and α and β are the params of the beta distr.

(a) Suppose the prior distribution for θ (the recovery rate) is Beta(2,2). Calculate the posterior distribution after observing the results of the experiment.

I used chat GPT to here to figure out the steps to calculate the posterior distr. by hand, ie formulas for α_{post} & β_{post}

Step 1: Prior Distr. of X is...

Prior: $Beta(\alpha_{prior}, \beta_{prior}) = Beta(2,2)$

Step 2: Likelihood - follows a binomial distr.

Likelihood $(K | n, \theta)$

Likelihood $(K|n,\theta) = \binom{n}{k} \theta^{k}(1-\theta)^{n-k}$

$K = 30$ (# of recoveries)

$n = 100$ (# of all patients)

θ = the recovery rate we are estimating

Step 3: Posterior Distribution

$\alpha_{post} = \alpha_{prior} + K = 2 + 30 = 32$

$\beta_{post} = \beta_{prior} + (n-K) = 2 + (100-30) = 72$

Posterior Distribution: Beta(32,72)

(b) Based on the posterior distribution, provide an estimate for θ.

To do this, we can use the mean of the Beta Distr. as a point estimate of X.

mean of Beta Distr.: $E[X] = \dfrac{\alpha}{\alpha+\beta} = \dfrac{32}{32+72} = \dfrac{32}{104} = 0.308$

$\boxed{E[X] = 0.308}$, meaning that after observing the experiment, the estimated recovery rate is about 30.8%.

Courtney Hodge

Problem 9: Conjugate Priors and Posterior Distributions

(c) The conjugate prior ensures that the posterior remains a Beta distribution, making the process efficient and computationally simple. Instead of performing complex calculations to derive the posterior distribution, we can just update the parameters of the prior.

Extra Credit: Non-Informative Priors

next page →

Courtney Hodge

Extra Credit

- The data
  1) comes from a normal dist. $N(\mu, \sigma^2)$, where $\sigma = 5$
  2) uses a non-informative conjugate prior μ, where $\mu \sim N(0, 100^2)$, meaning $\mu_{prior} = 0$ & $\sigma^2_{prior} = 10{,}000$
  3) is drawn independently from $N(\mu, 5^2)$

(a) Suppose we use a vague (non-informative) conjugate prior for μ, ie a normal distribution $N(0, 100^2)$. Compute the posterior distribution for μ after observing the data.

Likelihood $(X|\mu) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2\right)$

Prior distribution

$$P(\mu) = \frac{1}{\sqrt{2\pi\, 10{,}000}} \exp\left(-\frac{\mu^2}{2\times 10000}\right), \text{ since}$$

$\mu \sim N(0, 100^2)$

Since the posterior is proportional to the product of the likelihood and the Prior, and both the likelihood & prior are normally distributed, ~~then~~ and the conjugate prior for the mean of a normal distribution is also normal, the posterior will also be normal.

Posterior Mean formula: $\mu_{post} = \dfrac{\frac{n\bar{X}}{\sigma^2} + \frac{\mu_{prior}}{\sigma^2_{prior}}}{n/\sigma^2 + 1/\sigma^2_{prior}}$

posterior variance formula: $\sigma^2_{post} = \dfrac{1}{n/\sigma^2 + 1/\sigma^2_{prior}}$

where n = # of data points, $\bar{X}$ = the sample mean, $\sigma^2 = 25$ (known variance of the likelihood), $\mu_{prior} = 0$, $\sigma^2_{prior} = 10{,}000$

Courtney Hodge

(a) continued...

Posterior mean: $\mu_{post} = \dfrac{\frac{n\bar{x}}{25} + \frac{0}{10000}}{\frac{n}{25} + \frac{1}{10000}} = \dfrac{n\bar{x}}{n + 25/10000} =$

$$\mu_{post} = \frac{n\bar{x}}{n + 0.0025}$$

posterior variance $\sigma^2_{post} = \dfrac{1}{n/25 + 1/10000} = \dfrac{1}{n/25 + 0.001}$

Posterior Dist for μ:

$$\mu | X \sim N\left(\frac{n\bar{x}}{n + 0.0025}, \frac{1}{n/25 + 0.0001}\right)$$

(b) Now considering a Jeffreys prior... compute the posterior distribution for μ in this case.

$P(\mu) = 1$, for $-\infty < \mu < \infty$

$X_i \sim N(\mu, \sigma^2)$, where $\sigma^2 = 25$

The Jeffrey's prior is flat, meaning $P(\mu) = 1$

$$L(X|\mu) \propto \exp\left(\frac{-1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2\right)$$

Posterior Mean: $\mu_{post} = \bar{x}$

posterior Variance: $\sigma^2_{post} = \dfrac{\sigma^2}{n} = \dfrac{25}{n}$

$$\boxed{\mu | X \sim N\left(\bar{x}, \frac{25}{n}\right)}$$

(c) Compare the posterior distributions from the vague conjugate prior and the Jeffrey's prior. How do they differ and what might cause this difference?

The conjugate prior and Jeffreys prior differ in how they handle the prior information. The conjugate prior incorporates a normal distr. centered at 0, affecting the posterior distr.'s mean and variance. The Jeffrey's prior is flat and doesn't incorporate prior info to estimate μ or σ^2. This is b/c the Jeffreys prior only relies on the data.

Courtney Hodge

(d) Discuss the pros and cons of using non-informative priors in general, and how choosing different types of non-informative priors might affect your inferences.

Non-informative priors allows the data to drive the inference (Jeffreys priors), can simplify modeling (Jeffreys & conjugate), and can limit the introduction of bias. When we're missing information in our problem, use non-informative priors. Non-informative priors, <u>however</u>, can lead to misleading results and can become technically problematic with small datasets, for example.