Mid-term examination for course of Fundamentals of Statistical Machine Learning (20181115).

During the exam you may consult our text book "Information Theory, Inference, and Learning Algorithms" (David J.C. MacKay), and your own handwritten notes.

You may not consult anything else, including calculators. Use symbols for irrational numbers, but simplify rational expressions. So in your answers: $\lg 4 \to 2$, but leave $lg3$ as $lg3$. For clarity, use $\lg$ or $\log_2$ for logarithm base 2 and $\ln$ for natural logarithm.

You will probably find some problems much easier than others. So please read all the problems first (and write your name on each sheet of paper). If you have a question about the wording of a problem *during the first 15 minutes of the test*, I will be happy to try to clarify.

## This pdf list solutions along with the questions.

Your Name: ____________________

| value of $x$ | Probability |
|---|---|
| 0 | ½ |
| 1 | ¼ |
| 3 | ¼ |

Table 1: Probability distribution $f(x)$.

1. Let $f(x)$ denote the probability distribution of the integers $\{0, 1, 3\}$ shown in the table above. Let $G(n)$ be defined as the probability distibution over $\{0, 1, ..., 3n\}$ obtained by summing the value of $n$ independent samples from $f(x)$. So, for example, $G(2)$ takes a value of $6$ with probability $1/16$.

   What is the mean and variance of $G(n)$?

**Solution:** This problem utilizing the "expectation of a sum, is the sum of the expection" pattern for sums of independent variables.

The mean and variance of a single sample from $f(x)$, is:

$$\mu_1 \equiv \quad E[f(x)] = \quad 0 * 1/2 + 1 * 1/4 + 3 * 1/4 = 1.0$$

$$\sigma_1^2 \equiv \quad E[(f(x) - \mu_1)^2] = \quad (0-1)^2 1/2 + (1-1)^2 1/4 + (1-3)^2 1/4 = 1/2 + 4 * 1/4 = 1.5$$

Since the problem states that $G(n)$ is the sum of independent samples from $f(x)$, the mean and variance of $G(n)$ are $n\mu_1 = n$ and $n\sigma_1^2 = 1.5n$, respectively.

Your Name: ______________________________

2. (a) Let $f(x) = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots$ represent a probability distribution over the natural numbers $1, \ldots$. In other words, $\text{Prob}[f(x) = i] \equiv 2^{-i}$. What is the entropy of $f(x)$?

   (b) Similarly, let $g(x) = \frac{1}{3}, \frac{1}{3}, \frac{1}{9}, \frac{1}{9}, \frac{1}{27}, \frac{1}{27}, \ldots$ represent another probability distribution over the natural numbers. So, for example, $g(9)$ and $g(10)$ are both equal to $3^{-5}$. What is the entropy of $g(x)$?

**Solution:** This problem addresses the class partion property of entropy; imagine a discrete probability distribution $B$ over some classes with an entropy of $H(B)$ and let $B'$ denote a variant of $B$ in which one class with probability $p$ has been divided into a distribution over subclasses with entropy $H(C)$, then $H(B') \equiv pH(C)$. Above, $f(x)$ and $g(x)$ are defined by recursively dividing one class into equal halves and thirds respectively. Thus:

$$H(f) = H(1/2, 1/2) + 1/2 * H(1/2, 1/2) + 1/4 * H(1/2, 1/2) + \ldots = 1 + 1/2 + 1/4 + \ldots = 2$$

$$\begin{aligned} H(g) &= H(1/3, 1/3, 1/3) + 1/3 H(1/3, 1/3, 1/3) + 1/9 H(1/3, 1/3, 1/3) + \ldots \\ &= 1 \lg 3 + 1/3 \lg 3 + 1/9 \lg 3 + \ldots = \sum_{i=0}^{\infty} 3^{-i} \lg 3 = 3/2 \lg 3 \end{aligned}$$

Alternatively the definition of entropy can be applied directly to $f(x)$ and $g(x)$, leading to

$$\sum_{i=1}^{\infty} \left(2^{-i} \lg(2^i)\right) = \sum_{i=1}^{\infty} i\, 2^{-i} \equiv \sum_{i=0}^{\infty} 2^{-i} = 2$$

$$\sum_{i=1}^{\infty} \left(2 * 3^{-i} \lg(3^i)\right) = 2 * \sum_{i=1}^{\infty} \left(3^{-i} * i \lg 3\right) = 2 \lg 3 \sum_{i=1}^{\infty} i 3^{-i} = 2 \lg 3 * 3/4 = 3/2 \lg 3$$

Your Name: ______________________________

3. Let $D_1$ and $D_2$ denote the result of two independent rolls of a fair 4-sided die, and $S$ the sum of those two rolls (ranging from 2 to 8). More formally, $D_1$ and $D_2$ are independent random variables from a uniform distribution over $\{1,2,3,4\}$, and $S \equiv D_1 + D_2$.

   What is the mutual information between $D_1$ and $S$?

**Solution:** Apply identity $I(D_1, S) = H(S) - H(S|D_1)$. Straightforward computations for $H(S)$ are shown here:

| value | $p$ | $-\lg p$ | $-p \lg p$ |
|---|---|---|---|
| 2 | $1/16$ | 4 | $4/16$ |
| 3 | $2/16$ | 3 | $6/16$ |
| 4 | $3/16$ | $4 - \lg 3$ | $3/16(4 - \lg 3)$ |
| 5 | $4/16$ | 2 | $8/16$ |
| 6 | $3/16$ | $4 - \lg 3$ | $3/16(4 - \lg 3)$ |
| 7 | $2/16$ | 3 | $6/16$ |
| 8 | $1/16$ | 4 | $4/16$ |

$$\begin{aligned} H(S) &= 2 * (4/16 + 6/16 + 3/16(4 - \lg 3)) + 8/16 \\ &= 2 * (\quad 10/16 + \quad 12/16 - 3/16 \lg 3) + 8/16 \\ &= 10/8 + 12/8 - 3/8 \lg 3 + 4/8 \\ &= 26/8 - 3/8 \lg 3 \\ &= 3.25 - 3/8 \lg 3 \end{aligned}$$

Since $S$ is defined as $D_1 + D_2$, the remaining uncertainty of $S$ after knowing $D_1$ is the entropy of $D_2$ which is $\lg 4 = 2$. $I(D_1, S) \equiv H(S) - H(S|D_1) = 3.25 - 3/8 \lg 3 - 2 = 1.25 - 3/8 \lg 3$ ($\approx 0.6556$).

**Alternatively**, several students used the more convenient: $I(D_1, S) = H(D_1) - H(D_1|S)$.
$H(D_1|S)$ is:

| S=val | Prob S=val | #possible $D_1$ | $H(D_1 \mid S = \text{val})$ |
|---|---|---|---|
| 2 | $1/16$ | 1 | 0 |
| 3 | $2/16$ | 2 | 1 |
| 4 | $3/16$ | 3 | $\lg 3$ |
| 5 | $4/16$ | 4 | 2 |
| 6 | $3/16$ | 3 | $\lg 3$ |
| 7 | $2/16$ | 2 | 1 |
| 8 | $1/16$ | 1 | 0 |

$$H(D_1|S) = 4/16 \times 1 + 6/16 \times \lg 3 + 4/16 \times 2 = 12/16 + 6/16 \times \lg 3 = 3/4 + 3/8 \lg 3$$

$$I(D_1|S) = H(D_1) - H(D_1|S) = 2 - 3/4 + 3/8 \lg 3 = 5/4 - 3/8 \lg 3 (\approx 0.6556)$$

Your Name: ____________________

4. There are three boxes of coins. Box S containing 2 silver coins, Box M containing 2 gold and 2 silver coins, and Box G containing 3 gold coins. One of the three boxes is selected at random (each box having probability $1/3$ of being selected), and then a coin is drawn from the selected box. The drawn coin happens to be silver.

   What is the probability that the selected box is Box S?

   Is the probability still $1/3$? If not, how could drawing a coin after selecting a box affect which box was selected?

   (For this problem only, you may write your answer in Chinese if you prefer, 但不要用草書體字.)

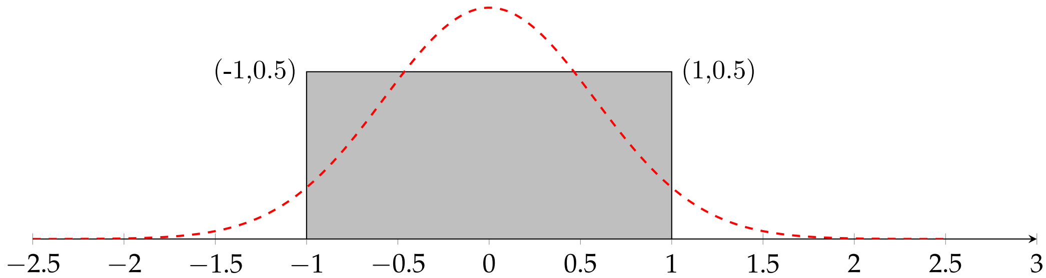
   **Solution: Q:** What is the probability that the selected box is Box S?

   **A:** Typical application of Bayes law. With uniform priors the posterior probability is proporitional to the likelihood. The probability of silver given S, M and G respectively is 1, $1/2$ and zero respectively. Normalizing to sum to one, the probabilities of S, M and G respectively are $2/3$, $1/3$ and zero.

   **Q:** Is the probability still $1/3$? If not, how could drawing a coin after selecting a box affect which box was selected?

   **A:** 老師你不要亂講! Drawing a coin does not affect which box was selected, what it does is *give us information about which box was selected.*

Your Name: ______________________________




5. Best fit of a normal distribution to a rectangular one.

   Imagine that numerical data are generated by taking independent samples from a distribution $U(x)$ uniform over (-1,1); like the gray rectangle shown above.

   $$\text{Probability Density}[x] \equiv \begin{cases} 0.5, & \text{if } -1 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

   Assume we are forced to model the data with a single normal distribution $N(0, \sigma^2)$, like the dotted curve in the figure at top.

   Question: What value of $\sigma^2$ maximizes the likelihood of data generated by $U(x)$ given the model $N(0, \sigma^2)$? (Give precise mathmetical justification for your answer).

   **Solution:** The data will consist of samples drawn independently from uniform [-1,1], so the likelihood of the data will be product of the likelihood of each data point. Therefore we want to maximize the product integral:

   $$\text{max over } \sigma \prod_{x=-1}^{+1} \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} \implies \text{max over } \sigma \int_{x=-1}^{+1} \left(\frac{-x^2}{2\sigma^2} - \ln\sigma\right)$$

   Where on the right we have taken the natural logarithm and kept only terms dependent on $\sigma$. Solving the definite integral gives:

   $$\int_{x=-1}^{+1} \left(\frac{-x^2}{2\sigma^2} - \ln\sigma\right) = \frac{-x^3}{6\sigma^2} - x\ln\sigma \Big|_{x=-1}^{+1} = \frac{-1}{3\sigma^2} - 2\ln\sigma$$

   At a maximum:

   $$0 = \frac{d}{d\sigma}\left\{\frac{-1}{3\sigma^2} - 2\ln\sigma\right\} = \frac{2}{3\sigma^3} - \frac{2}{\sigma} \implies 3\sigma^3 = \sigma;\ \sigma^2 = 1/3.$$

If extra space needed, you may use the following page as well.

Your Name: ______________________________

Continuation of question from previous page about best fit of a normal distribution to a rectangular one.

Your Name: ____________________

6. Data $X = x_1, \ldots$ is generated by a mixture model with components M1 and M2. When generating a data point; first one of the components is selected using random numbers — selecting M1 with probability 1/4 (and therefore M2 with probability 3/4).

   M1 is a normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$.
   M2 is a normal distribution with mean $\mu = 3$ and variance $\sigma^2 = 2$.

   The first point generated is observed to have value of one (i.e. $x_1 = 1$).

   What is the probability that $x_1$ was generated by component M1?

**Solution:** Using the odds ratio version of Bayes law

$$\text{Posterior odds} = \frac{\text{Prob}[\text{M}_1|\text{D}]}{\text{Prob}[\text{M}_2|D]} \equiv \frac{\text{Prob}[\text{M}_1]}{\text{Prob}[\text{M}_2]} \, \frac{\text{Prob}[\text{D}|\text{M}_1]}{\text{Prob}[\text{D}|\text{M}_2]}$$

$$\text{Prob}[\text{D}|\text{M}] \propto \sigma^{-1} e^{-z^2/2}, \quad z^2 \stackrel{\text{def}}{=} \frac{(\mu - x)^2}{\sigma^2}$$

From the problem statement, $\frac{\text{Prob}[\text{M}_1]}{\text{Prob}[\text{M}_2]} = \frac{1/4}{3/4} = 1/3$. The other numbers given yield:

$$\sigma_1 = 1, \; z_1^2 = \frac{(1-0)^2}{1} = 1\,; \qquad \sigma_2 = \sqrt{2}, \; z_2^2 = \frac{(1-3)^2}{2} = 2$$

$$\frac{\text{Prob}[\text{D}|\text{M}_1]}{\text{Prob}[\text{D}|\text{M}_2]} = \frac{1^{-1}\, e^{-1/2}}{\sqrt{2}^{-1} e^{-2/2}} = \frac{\sqrt{2}\, e^{2/2}}{e^{1/2}} = \sqrt{2e} \quad \Longrightarrow \quad \frac{\text{Prob}[\text{M}_1|\text{D}]}{\text{Prob}[\text{M}_2|D]} = \frac{1}{3}\sqrt{2e} = \frac{\sqrt{2e}}{3}.$$

In general, when the odds ratio of a proposition and its negation are $x : y$, then the probability of the proposition is $\frac{x}{x+y}$.

$$\frac{\text{Prob}[\text{M}_1|\text{D}]}{\text{Prob}[\text{M}_2|D]} = \frac{\sqrt{2e}}{3} \Longrightarrow \text{Prob}[\text{M}_1|\text{D}] = \frac{\sqrt{2e}}{\sqrt{2e}+3} \; (\approx 0.4373)$$