

CompSci 527 Midterm Exam Sample— Sample Solution

1. The following table shows the joint probability $p(x, y)$ for random variables X and Y . The variable X can take values in $\{1, 2, 3\}$ and the variable Y can take values in $\{0, 1\}$. Fill the empty cells in the tables below. Represent values as irreducible fractions of integers ($2/3$ but not $4/6$) or as decimals (0.4). The decimal representation of $2/3$ is $0.\overline{6}$. If in doubt about your numbers, show your calculations *clearly* for partial credit.

$p(x, y)$	$x = 1$	2	3
$y = 0$	0.1	0.1	0.2
1	0.3	0.0	0.3

	$p(y)$
$y = 0$	0.4
1	0.6

	$x = 1$	2	3
$p(x)$	0.4	0.1	0.5

$$E[X] = \boxed{2.1} \quad \sigma_X^2 = \boxed{0.89}$$

$$E[Y] = \boxed{0.6} \quad \sigma_Y^2 = \boxed{0.24}$$

$p(x y)$	$x = 1$	2	3
$y = 0$	0.25	0.25	0.50
1	0.50	0.00	0.50

$p(y x)$	$x = 1$	2	3
$y = 0$	0.25	1.00	0.40
1	0.75	0.00	0.60

Answer: *The calculations for the means and variances are as follows:*

$$\begin{aligned}
 E[X] &= 0.4 \times 1 + 0.1 \times 2 + 0.5 \times 3 = 0.4 + 0.2 + 1.5 = 2.1 \\
 E[Y] &= 0.4 \times 0 + 0.6 \times 1 = 0.6 \\
 \sigma_X^2 &= 0.4 \times (1 - 2.1)^2 + 0.1 \times (2 - 2.1)^2 + 0.5 \times (3 - 2.1)^2 \\
 &= 0.4 \times 1.21 + 0.1 \times 0.01 + 0.5 \times 0.81 \\
 &= 0.484 + 0.001 + 0.405 = 0.89 \\
 \sigma_Y^2 &= 0.4 \times (0 - 0.6)^2 + 0.6 \times (1 - 0.6)^2 \\
 &= 0.4 \times 0.36 + 0.6 \times 0.16 \\
 &= 0.144 + 0.096 = 0.24
 \end{aligned}$$

2. In the table below, write the joint probability $p(x, y)$ of two *independent* random variables X and Y that have the same marginal distributions $p(x)$ and $p(y)$ as in the previous problem.

$p(x, y)$	$x = 1$	2	3
$y = 0$	0.16	0.04	0.20
1	0.24	0.06	0.30

Answer: *To obtain this table, just multiply the two marginals from the previous problem.*

3. A Bernoulli random variable X with parameter λ has universe $\{0, 1\}$ and distribution $p(0) = \mathbb{P}[X = 0] = 1 - \lambda$ and $p(1) = \mathbb{P}[X = 1] = \lambda$. Show that the mean and variance of X are $m_X = \lambda$ and $\sigma_X^2 = \lambda(1 - \lambda)$.

Answer:

$$\begin{aligned}
 m_X &= E[X] = 0 \times \mathbb{P}[X = 0] + 1 \times \mathbb{P}[X = 1] = \mathbb{P}[X = 1] = \lambda \\
 \sigma_X^2 &= E[(X - m_X)^2] = (0 - \lambda)^2 \times \mathbb{P}[X = 0] + (1 - \lambda)^2 \times \mathbb{P}[X = 1] \\
 &= (0 - \lambda)^2(1 - \lambda) + (1 - \lambda)^2\lambda = \lambda^2(1 - \lambda) + (1 - \lambda)^2\lambda \\
 &= (1 - \lambda)[\lambda^2 + (1 - \lambda)\lambda] = (1 - \lambda)\lambda
 \end{aligned}$$

4. The Dirichlet distribution over $K = 2$ continuous values λ_1, λ_2 and with parameters $\alpha_1 = \alpha_2 = 2$ has density

$$p(\lambda_1, \lambda_2) = c\lambda_1\lambda_2 \quad \text{wherever} \quad \lambda_1 + \lambda_2 = 1 \quad \text{and} \quad \lambda_1 \geq 0, \lambda_2 \geq 0,$$

and is zero elsewhere. What is the value of c ? Show your reasoning. [Hint: Compute c from first principles rather than trying to remember a formula. What part of the (λ_1, λ_2) plane satisfies the constraints in the definition of p ?]

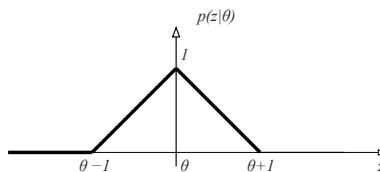
Answer: The density must integrate to 1, so

$$c^{-1} = \int \int_{\lambda_1 + \lambda_2 = 1, \lambda_1 \geq 0, \lambda_2 \geq 0} \lambda_1 \lambda_2 d\lambda_1 d\lambda_2 = \int_0^1 (1 - \lambda_1) \lambda_1 d\lambda_1$$

because the constraints reduce the domain to a single line segment on the (λ_1, λ_2) plane. Therefore,

$$c^{-1} = \int_0^1 (1 - x)x dx = \int_0^1 x dx - \int_0^1 x^2 dx = \frac{x^2}{2} \Big|_0^1 - \frac{x^3}{3} \Big|_0^1 = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}.$$

5. The triangle distribution with parameter θ is a continuous probability distribution defined as follows:

$$p(z|\theta) = \begin{cases} 1 + z - \theta & \text{for } \theta - 1 \leq z \leq \theta \\ 1 - z + \theta & \text{for } \theta \leq z \leq \theta + 1 \\ 0 & \text{elsewhere} \end{cases}$$


and is shown in the figure. You are given a training set with two samples

$$z_1 = 0.3 \quad \text{and} \quad z_2 = 0.5$$

drawn out of the triangle distribution $p(z|\theta)$, where the unknown parameter θ is a random variable that has uniform distribution between 0 and 0.3.

- (a) Find the Maximum Likelihood (ML) estimate $\hat{\theta}_{ML}$ of θ and explain your answer. [Hint: A drawing and symmetry considerations may help. Brute force is not recommended.]

Answer: The question is where to place the center θ of a single triangle distribution so that the product

$$L(z_1, z_2|\theta) = p(z_1|\theta)p(z_2|\theta)$$

is as large as possible. By symmetry, and since p is monotonic and linear away from its maximum, the only reasonable options are $\theta = z_1$ or $\theta = z_2$ or $\theta = (z_1 + z_2)/2$. The first two choices place one point at the maximum of $p(\cdot|\theta)$, with a likelihood value of 1, and the other point 0.2 units away, with a likelihood value of $1 - 0.2 = 0.8$, so that $L = 1 \times 0.8 = 0.8$ with either of these choices. The third choice places both samples 0.1 units from the maximum, so that $L = (1 - 0.1)^2 = 0.9^2 = 0.81$, which is greater. So

$$\hat{\theta}_{ML} = \frac{z_1 + z_2}{2} = \frac{0.3 + 0.5}{2} = 0.4.$$

An more detailed version of this reasoning would be to observe that the ML criterion is piecewise quadratic in θ , increasing to the left of z_1 , and decreasing to the right of z_2 . So the maximum of L is between z_1 and z_2 . By symmetry, the maximum is therefore either indifferently at $\theta = z_1$ or $\theta = z_2$, or it is at the midpoint $\theta = (z_1 + z_2)/2$ between them. Checking the value of L at these three points yields the answer above.

- (b) Find the Maximum a Posteriori (MAP) estimate $\hat{\theta}_{MAP}$ of θ and explain your answer. [Hint: Do you even have a choice?]

Answer: The prior on θ is uninformative between 0 and 0.3, and rules out any values outside that interval. The ML criterion—and therefore the MAP criterion—decreases to the left of $z_1 = 0.3$, so the MAP solution is

$$\hat{\theta}_{MAP} = 0.3.$$

6. A company makes pencils, and 20 percent of its pencils are broken. A quality check reports all the good pencils as such, but reports 10 percent of the broken ones as good. Use the following event symbols in your formulas (you may not need all of them):

B : a pencil is broken
 G : a pencil is good
 P : a pencil passes the check
 F : a pencil fails the check

- (a) What is the probability that a pencil that passes the check is good? Show your reasoning. You may leave your answer as a ratio of two (but no more) decimals (example: 0.42/0.65).

Answer:

$$\begin{aligned}
 P(B) &= 0.2 \\
 P(G) &= 1 - P(B) = 0.8 \\
 P(P|G) &= 1 \\
 P(P|B) &= 0.1 \\
 P(P) &= P(P|G)P(G) + P(P|B)P(B) = 1 \times 0.8 + 0.1 \times 0.2 = 0.82 \\
 P(G|P) &= \frac{P(P|G)P(G)}{P(P)} = \frac{1 \times 0.8}{0.82} = \frac{1 \times 0.8}{0.82} [\approx 0.9756]
 \end{aligned}$$

- (b) If the company sells all pencils that pass the check, what is the probability that it sells a bad one? Show your reasoning. You may leave your answer as a ratio of two (but no more) decimals (example: 0.42/0.65).

Answer:

$$P(B|P) = 1 - P(G|P) = 1 - \frac{0.8}{0.82} = \frac{0.82 - 0.8}{0.82} = \frac{0.02}{0.82} [\approx 0.0244]$$

7. A world quantity w with standard uniform prior

$$p(w) = \begin{cases} 1 & \text{for } 0 \leq w \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

is observed through scalar measurements with a standard normal likelihood

$$p(x|w) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-w)^2}{2}}.$$

- (a) Write an explicit formula for the Bayes regressor

$$\hat{w} = f(x).$$

“Explicit” here entails that your final formula cannot contain the symbols $\int$, p , or w . You may of course use these symbols in your derivation. [Hint: Once you have a (correct!) formula for the joint density of x and w , consider the three cases $x < 0$ and $0 \leq x \leq 1$ and $x > 1$ separately. A drawing may help.]

Answer: From Bayes’s theorem, the posterior distribution of w given the data x is

$$p(w|x) = \frac{p(x|w)p(w)}{p(x)}$$

where the denominator is independent of w and can therefore be ignored in the Bayes regressor

$$\hat{w} = f(x) = \arg \max_w p(w|x) = \arg \max_w \frac{p(x|w)p(w)}{p(x)} = \arg \max_w p(x|w)p(w) = \arg \max_w p(x, w).$$

The joint distribution $p(x, w)$ is

$$p(x, w) = p(x|w)p(w) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-w)^2}{2}} & \text{for } 0 \leq w \leq 1 \\ 0 & \text{elsewhere} \end{cases}.$$

While $p(x, w)$ is zero for w outside the unit interval, it is nonzero for all values of x , so we consider the three cases $x < 0$, $0 \leq x \leq 1$, and $x > 1$ separately when maximizing this expression over w for a fixed x .

When $x < 0$, the joint distribution $p(x, w)$, when viewed as a function of w is monotonically decreasing for $0 \leq w \leq 1$, because the peak of the likelihood $p(x|w)$, viewed as a function of w , is to the left of $w = 0$. So

$$\hat{w} = f(x) = 0 \quad \text{when} \quad x < 0.$$

Similar reasoning leads to

$$\hat{w} = f(x) = 1 \quad \text{when} \quad x > 0.$$

When $0 \leq x \leq 1$, the maximum of the likelihood is at $w = x$:

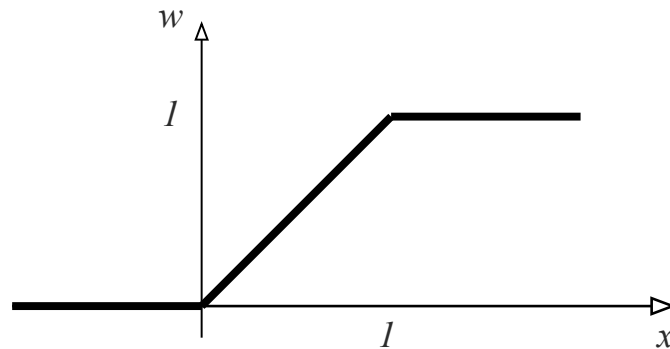
$$\hat{w} = f(x) = x \quad \text{when} \quad 0 \leq x \leq 1.$$

In summary,

$$\hat{w} = f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } 0 \leq x \leq 1 \\ 1 & \text{for } x > 1 \end{cases}$$

- (b) Sketch your regressor function $f(x)$, and label the axes accurately.

Answer:



8. You are asked to design an inference system to track the puck in an ice hockey game from a video recording of the game. Assume that you only need to track the puck when it is visible. Suppose that it is straightforward to model the imaging process as a function $\mathbf{y} = g(\mathbf{w})$ that computes the ideal image $\mathbf{y}$ of the puck from the 3D coordinates $\mathbf{w}$ of the puck's centroid in the world. The imaging process is noisy, so the real image is

$$\mathbf{x} = \mathbf{y} + \mathbf{n}$$

where $\mathbf{n}$ is random noise of known probability distribution.

- (a) Is your inference system a classifier or a regressor? Why?

Answer: A regressor, because it needs to estimate the value $\hat{\mathbf{w}}$ of a quantity that varies continuously, that is, the position of the puck in the world.

- (b) Would you design a generative or a discriminative inference system? Justify your answer.

Answer: A generative system, because the deterministic function g and distribution of the noise term $\mathbf{n}$ are equivalent to specifying the likelihood $p(\mathbf{x}|\mathbf{w})$. Modeling the posterior $p(\mathbf{w}|\mathbf{x})$ directly would be challenging with the information at hand.

- (c) What else do you need to know or estimate to complete your design, in addition to the function g and the probability distribution of $\mathbf{n}$? Why?

Answer: I need to know the prior $p(\mathbf{w})$ in order to compute the posterior (or joint) distribution from the likelihood through Bayes's theorem.

- (d) What does a training set for this problem look like? Describe a training sample, and state what annotations (human-entered information) are needed for it. [No need to state how training samples are collected.]

Answer: Each data point is a video frame with an image of the puck delineated by hand (the annotation) and the corresponding position of the puck in the world.