HW0 (4 problems)

Jamie Morgenstern *CSE515*

Basic probability.

1. *[2 points]* For three discrete random variables x, y, and $z \in \Omega$ with a joint probability distribution $p(x,y,z)$, we say x is **conditionally independent** of y given z if and only if $p(x,y|z) = p(x|z)p(y|z)$, where $p(x|z) = \frac{p(x,z)}{p(z)} = \frac{\sum_{y' \in \Omega} p(x,y',z)}{\sum_{x' \in \Omega, y' \in \Omega} p(x',y',z)}$ is the conditional probability distribution of x given z, and $p(x,y|z)$ and $p(y|z)$ are defined similarly. We use the notation x–z–y to denote that x and y are independent conditioned on z. We use x–()–(y,z) to denote that x is independent of the pair of random variables (y,z) (that is $p(x,y,z) = p(x)p(y,z)$). Here is an example of a formal proof I expect from your solutions. I will prove that x–()–(y,z) implies x–()–z, that is x independent of (y,z) implies x independent of z. This sounds obviously true, and it is true. The point is to teach you how to write the proof.

   Proof. By the assumption x–()–(y,z), we have $p(x,y,z) = p(x)p(y,z)$. This implies that

$$\begin{aligned} p(x,z) &= \sum_{y' \in \Omega} p(x,y',z) \\ &= \sum_{y' \in \Omega} p(x)p(y',z) \\ &= p(x)p(z)\ , \end{aligned} \tag{1}$$

   where in the last equality we used the fact that $p(z) = \sum_{x' \in \Omega, y' \in \Omega} p(x',y',z) = \sum_{x' \in \Omega, y' \in \Omega} p(x')p(y',z) = \sum_{y' \in \Omega} p(y',z)$ as probability measures sum to one and $\sum_{x'} p(x') = 1$. By Eq.(1) we have the desired claim: x–()–z. □

   This is a little excessive. You don't have to write all the details like I did in this example. It was just to make sure everyone can follow every step. In your solutions, you can skip some obvious steps but have to write enough details to make sure that we know what you are doing.

   Now, prove each of the following properties.

   (a) x–()–(y,z) implies x–z–y
   (b) x–z–(y,w) implies x–z–y
   (c) x–z–(y,w) and y–z–w implies (x,w)–z–y

2. *[2 points]* There are two coins, one is a fair coin and the other is biased. The outcome of tossing a fair coin is a head (H) with probability half. The outcome of tossing a biased coin is a head (H) with probability 3/4. We are given a coin at random (equal probability of getting a fair or a biased coin), and want to test whether the coin is biased or not based on n coin tosses. We toss the coin 5 times independently and get (H,H,T,T,H). Let $x \in \{\text{fair,biased}\}$ denote the random variable for the choice of the coin and let $y = (H,H,T,T,H) \in \{H,T\}^5$ denote the observation from 5 tosses of the coin.

(a) What is the probability that the coin is a biased coin given the observations? In other words, compute $\mathbb{P}(x = \text{biased}|y = (H, H, T, T, H))$.

(b) What is your **maximum a posteriori (MAP) estimate**? A MAP estimate is defined as the most likely estimate of the hidden variable x given the observations: $\arg\max_{x\in\{\text{fair},\text{biased}\}} \mathbb{P}(x|y)$.

(c) We want to show that the count of the number of heads is a sufficient statistic. Prove that the MAP estimate does not depend on the order of the outcome.

Basic linear algebra.

3. *[2 points]* An $n \times n$ dimensional symmetrix matrix A is **positive definite** if $x^T Ax > 0$ for any vector $x \in \mathbb{R}^n$ and $x \neq 0$, where x^T is the transpose of the column vector x. It is **positive semidefinite** if $x^T Ax \geq 0$ for any x. Prove that if A has eigen values which are all positive, then A is positive definite.

   [Hint: A symmetric matrix A can be factorized by eigen decomposition as $A = Q\Lambda Q^T$. Q is a unitary matrix such that $QQ^T = Q^T Q = I$, where I is the n-dimensional identity matrix, and Λ is a diagonal matrix with the eigen values λ_i of matrix A in the diagonals. Then, we are left to show that if Λ is a diagonal matrix with strictly positive entries, then $y^T\Lambda y > 0$, where we changes variables by setting $y = Q^T x$.].

4. *[2 points]* Find a vector y^* in terms of A_{ij}'s, h_i's and x that minimizes a **quadratic function**

$$f(x, y) \;=\; \begin{bmatrix} x^\top & y^\top \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} h_1^\top & h_2^\top \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} ,$$

   where A_{22} is a symmetric and positive definite matrix. Here, $x \in \mathbb{R}^{d_x}$, $y \in \mathbb{R}^{d_y}$, $A_{11} \in \mathbb{R}^{d_x \times d_x}$, $A_{12} \in \mathbb{R}^{d_x \times d_y}$, $A_{21} \in \mathbb{R}^{d_y \times d_x}$, $A_{22} \in \mathbb{R}^{d_y \times d_y}$, $h_1 \in \mathbb{R}^{d_x}$, and $h_2 \in \mathbb{R}^{d_y}$. You need to show your derivation. [Note: the minimizer $y^*(x)$ is a linear function of x.]