# Home Assignment №1

**Due on** October 11, 2024, 23:59 PM

## Exercise 1

[5 points]. This problem reviews basic concepts from probability.

a) [1 point]. A biased die has the following probabilities of landing on each face:

| face | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| P(face) | .1 | .1 | .2 | .2 | .4 | 0 |

I win if the die shows even. What is the probability that I win? Is this better or worse than a fair die (i.e., a die with equal probabilities for each face)?

b) [1 point]. Recall that the expected value $\mathbb{E}[X]$ for a random variable $X$ is

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} p(X = x)\ x,$$

where $\mathcal{X}$ is the set of values $X$ may take on. Similarly, the expected value of any function $f$ of random variable $X$ is

$$\mathbb{E}[f(X)] = \sum_{x \in \mathcal{X}} p(X = x)\ f(x).$$

Now consider the function below, which we call the "indicator function"

$$\mathbb{I}[X = a] := \begin{cases} 1 & \text{if } X = a \\ 0 & \text{if } X \neq a \end{cases}.$$

Let $X$ be a random variable which takes on the values $3, 8$ or $9$ with probabilities $p_3$, $p_8$ and $p_9$ respectively. Calculate $\mathbb{E}[\mathbb{I}[X = 8]]$.

c) [2 points]. Recall the following definitions:

- Entropy: $H(X) = -\sum_{x\in\mathcal{X}} p(X=x)\log_2 p(X=x) \quad = -\mathbb{E}[\log_2 p(X)]$
- Joint entropy: $H(X,Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(X=x,Y=y)\log_2 p(X=x,Y=y) \quad = -\mathbb{E}[\log_2 p(X,Y)]$
- Conditional entropy: $H(Y|X) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(X=x,Y=y)\log_2 p(Y=y|X=x) \quad = -\mathbb{E}[\log_2 p(Y|X)]$
- Mutual information: $I(X;Y) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(X=x,Y=y)\log_2 \frac{p(X=x,Y=y)}{p(X=x)p(Y=y)}$

Using the definitions of the entropy, joint entropy, and conditional entropy, prove the following chain rule for the entropy:

$$H(X,Y) = H(Y) + H(X|Y).$$

d) [1 point]. Recall that two random variables $X$ and $Y$ are *independent* if

$$\text{for all } x\in\mathcal{X} \text{ and all } y\in\mathcal{Y}, \quad p(X=x,Y=y) \;=\; p(X=x)p(Y=y).$$

If variables $X$ and $Y$ are independent, is $I(X;Y)=0$? If yes, prove it. If no, give a counter example.

## Exercise 2

[4 points]. Given a training set $\mathcal{D} = \{(x^{(i)}, y^{(i)}), i = 1,\dots,M\}$, where $x^{(i)} \in \mathbb{R}^N$ and $y^{(i)} \in \{1,2,\dots,C\}$, derive the maximum likelihood estimates of the naive Bayes for real valued $x_j^{(i)}$ modeled with a Laplacian distribution, *i.e.*,

$$p(x_j|y=c) = \frac{1}{2\sigma_{j|c}}\exp\left(-\frac{|x_j - \mu_{j|c}|}{\sigma_{j|c}}\right).$$

## Exercise 3

[4 points]. Prove that in binary classification, the posterior of linear discriminant analysis, *i.e.*, $p(y=1|x;\varphi,\mu,\Sigma)$, is in the form of a sigmoid function

$$p(y=1|x;\theta) = \frac{1}{1+e^{-\theta^T x}}$$

where $\theta$ is a function of $\{\varphi,\mu,\Sigma\}$. <u>Hint:</u> remember to use the convention of letting $x_0 = 1$ that incorporates the bias term into the parameter vector $\theta$.

## Exercise 4

[2 points]. For an $N$-dimensional vector $x$, the multivariate Gaussian distribution takes the form

$$\mathcal{N}(x;\mu,\Sigma) = \frac{1}{\sqrt{(2\pi)^N|\Sigma|}} \exp\left\{-\frac{1}{2}\left(x-\mu\right)^T \Sigma^{-1}\left(x-\mu\right)\right\}. \tag{1}$$

We partition $x$ into two disjoint subsets $x_a$ and $x_b$. Without loss of generality, we can take $x_a$ to form the first $N_1$ elements of $x$, with $x_b$ comprising the remaining $N - N_1$ elements such that

$$x = \begin{bmatrix} x_a \\ x_b \end{bmatrix}, \tag{2}$$

$$\mu = \begin{bmatrix} \mu_a \\ \mu_b \end{bmatrix}, \tag{3}$$

and

$$\Sigma^{-1} = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}^{-1} = \begin{bmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{bmatrix}, \tag{4}$$

where $\Sigma_{ab}^T = \Sigma_{ba}$ and $\Lambda_{ab}^T = \Lambda_{ba}$. Prove that the conditional of a joint Gaussian distribution $x_b|x_a$ given by

$$p\left(x_b|x_a\right) = \frac{p\left(x_a, x_b; \mu, \Sigma\right)}{\int p\left(x_a, x_b; \mu, \Sigma\right) dx_b} \tag{5}$$

is also Gaussian.

<u>Hints:</u> You may derive the mean vector and the covariance matrix of $p(x_b|x_a)$ by comparing the coefficients of your expression with the following general form:

$$\frac{1}{2}z^T Az + b^T z + c = \frac{1}{2}\left(z + A^{-1}b\right)^T A\left(z + A^{-1}b\right) + c - \frac{1}{2}b^T A^{-1}b. \tag{6}$$

By the way, the method is called "completing the square".

Besides, you may find this more general result of block matrix inverse relating to Eq. (4) useful for interpreting your solution:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} M & -MBD^{-1} \\ -D^{-1}CM & D^{-1} + D^{-1}CMBD^{-1} \end{bmatrix} \tag{7}$$

where we have defined

$$M = (A - BD^{-1}C)^{-1}. \tag{8}$$