PPGEE2249 Aprendizado de Máquina
Assignment 1
Prof. Daniel Guerreiro e Silva

Thiago Tomás de Paula

November 8, 2025

▶ Question 1

X is a continuous-valued random variable with uniform density in $(-1, +1)$.

▷ Item a

Draw its probability density function (PDF). What is the area under the curve? Justify your answer.

▷ Item b

Draw its cumulative distribution function (CDF).

▷ Item c

Calculate the probability of the event $X \in (-0.2, 0.2)$.

▷ Item d

Calculate the expected value $\mathrm{E}[X]$, the second $\mathrm{E}[X^2]$ and the fourth moment $\mathrm{E}[X^4]$ of the random variable. Calculate its variance $\sigma^2[X]$, as well.

Let $p : \mathbb{R} \to \mathbb{R}$ denote the PDF of X and $c : \mathbb{R} \to \mathbb{R}$ its CDF. Since the random variable is of uniform density in $(-1, +1)$,

$$p(x) = \begin{cases} k, & |x| < 1 \\ 0, & |x| \geq 1 \end{cases} \tag{1}$$

for some constant $k > 0$. By definition, the area under the PDF curve is equal to the probability P of X assuming a value in the associated range (a, b), $P(a < X < b) = \int_a^b p(x)\ dx$. For $a = -\infty$ and $b = \infty$, $P(-\infty < X < \infty) = 1$, i.e., the probability of X assuming *any* value is 100%. We can use this result (in general, an identity) to find the value of k:

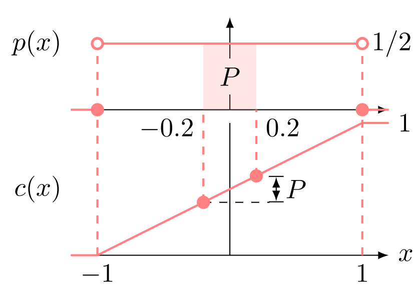
$$P(-\infty < X < \infty) = \int_{-\infty}^{\infty} p(x)\ dx = 1 \implies \int_{|x|\geq 1} p(x)\ dx + \int_{|x|<1} p(x)\ dx = 1 \tag{2}$$

$$\int_{|x|\geq 1} 0\ dx + \int_{|x|<1} k\ dx = 1 \tag{3}$$

$$2k = 1 \quad \therefore \quad k = \frac{1}{2}. \tag{4}$$

With k evaluated, we can find any P, e.g., $P(-0.2 < X < 0.2) = [0.2 - (-0.2)] \cdot (1/2) = 0.2$. A drawing of $p(x)$ and $c(x) = \int_{-\infty}^{x} p(z)\ dz$ is presented in fig. 1.

Figure 1: PDF and CDF of a uniform distribution in $(-1, 1)$.



Next, the variance and some of the moments of p is calculated. By definition, $\mathrm{E}[X^n] = \int_{\mathbb{R}} x^n p(x)\, dx$. For this question's distribution,

$$\mathrm{E}[X^n] = \int_{-1}^{1} x^n \cdot \frac{1}{2}\, dx = \frac{1}{2} \cdot \frac{x^{n+1}}{n+1}\bigg|_{x=-1}^{x=1} = \frac{1+(-1)^n}{2(n+1)} = \begin{cases} 0, & n \text{ odd} \\ 1/(n+1), & n \text{ even} \end{cases} \tag{5}$$

The above expression gives, for example, $\mathrm{E}[X] = 0$, $\mathrm{E}[X^2] = 1/3$, and $\mathrm{E}[X^4] = 1/5$. As for the variance, the definition $\sigma^2[X] = \mathrm{E}[(X - \mathrm{E}[X])^2] = \mathrm{E}[X^2] - \mathrm{E}[X]^2$ results in $\sigma^2[X] = 1/3$.

► Question 2

X is a discrete-valued random variable with uniform distribution over the set $\{-2, -1, 0, 1, 2\}$. Draw its probability mass function (PMF) and calculate $\mathrm{E}[X]$ and $\sigma^2[X]$.

Let $p : \mathbb{Z} \to \mathbb{R}$ denote the PMF of X. Since the random variable is uniform on $\{x \in \mathbb{Z} : |x| < 3\}$, we have

$$p(x) = \begin{cases} k, & |x| < 3 \\ 0, & |x| \geq 3 \end{cases} \tag{6}$$

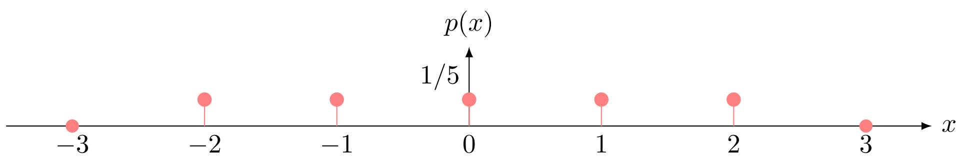
for some constant $k > 0$. Since the PMF must satisfy the identity $\sum_{x \in \mathbb{Z}} p(x) = 1$, we have

$$\sum_{x \in \mathbb{Z}} p(x) = \sum_{|x| \geq 3} p(x) + \sum_{|x| < 3} p(x) = 1 \tag{7}$$

$$0 + 5k = 1 \quad \therefore \quad k = \frac{1}{5} \tag{8}$$

The drawing in fig. 2 illustrates the PMF with k evaluated.

Figure 2: PMF of an uniform distribution in $\{-2, -1, 0, 1, 2\}$.



As for $\mathrm{E}[X]$ and $\sigma^2[X]$, we can use the definition for the n-th moment $\mathrm{E}[X^n] = \sum_{x \in \mathbb{Z}} x^n p(x)$ and the definition $\sigma^2[X] = \mathrm{E}[(X - \mathrm{E}[X])^2] = \mathrm{E}[X^2] - \mathrm{E}[X]^2$:

$$\mathrm{E}[X] = \sum_{|x|<3} \frac{x}{5} = \frac{1}{5}(-2 - 1 + 0 + 1 + 2) = 0 \tag{9}$$

$$\mathrm{E}[X^2] = \sum_{|x|<3} \frac{x^2}{5} = \frac{1}{5}[(-2)^2 + (-1)^2 + 0^2 + 1^2 + 2^2] = 2 \tag{10}$$

$$\sigma^2[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = 2 - 0 = 2 \tag{11}$$

► Question 3

Consider the normal random variables $X_1 \sim N(-2, 2)$, $X_2 \sim N(1, 4)$, with $\mathrm{cov}(X_1, X_2) = -0.8$. Calculate the joint pdf of $X = (X_1, X_2)^T$.

Let $X_i \sim N(\mu_i, \sigma_i^2)$, $1 \le i \le d$ be several univariate normal random variables. The joint PDF $p : \mathbb{R}^{d\times 1} \to \mathbb{R}$ of $\mathbf{x} = [x_1\ x_2\ \cdots\ x_d]^\mathrm{T}$ is the multivariate gaussian

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \times \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\mathrm{T}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right] \tag{12}$$

with $\boldsymbol{\mu} = [\mu_1\ \mu_2\ \cdots\ \mu_d]^\mathrm{T}$ and $\Sigma_{ij} = \sigma_{ij} = \mathrm{cov}(X_i, X_j)$ the covariance matrix of $\mathbf{x}$.

The question statement gives

$$\mu_1 = -2, \quad \sigma_1^2 = 2, \quad \mu_2 = 1, \quad \sigma_2^2 = 4, \quad \sigma_{12} = \sigma_{21} = -0.8 \tag{13}$$

which leads to

$$\boldsymbol{\mu} = \begin{bmatrix} -2 \\ 1 \end{bmatrix} \tag{14}$$

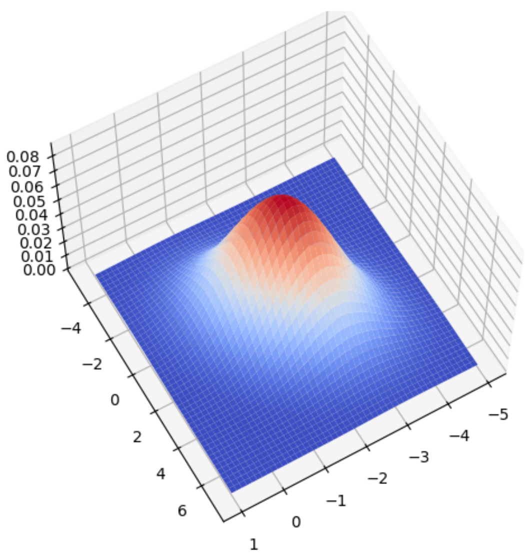
$$\boldsymbol{\Sigma} = \begin{bmatrix} 2 & -0.8 \\ -0.8 & 4 \end{bmatrix}, \quad |\boldsymbol{\Sigma}| = 8 - 0.36 = 7.64, \quad \boldsymbol{\Sigma}^{-1} = \frac{1}{7.64}\begin{bmatrix} 4 & 0.8 \\ 0.8 & 1 \end{bmatrix} \tag{15}$$

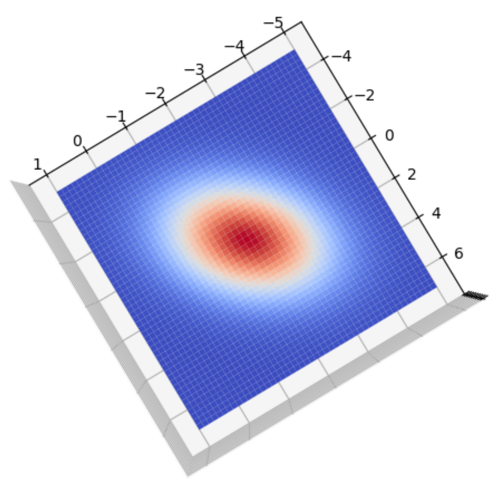
and, making all substitutions in (12),

$$p(\mathbf{x}) = \frac{1}{2\pi\sqrt{7.36}} \times \exp\left[-\frac{1}{2}\begin{bmatrix} x_1+2 \\ x_2-1 \end{bmatrix}^\mathrm{T} \frac{1}{7.64}\begin{bmatrix} 4 & 0.8 \\ 0.8 & 1 \end{bmatrix}\begin{bmatrix} x_1+2 \\ x_2-1 \end{bmatrix}\right]. \tag{16}$$

fig. 3 shows a plot of the multivariate distribution given by the expression above.

Figure 3: Joint PDF of $X_1 \sim N(-2, 2)$ and $X_2 \sim N(1, 4)$.





► Question 4

Assume you want to design a classifier which receives univariate data as input and must choose between class +1 or class −1. The data associated with class "+1" follow a gaussian density with mean 2 and unit variance. On the other hand, the data associated with class "−1" follow a uniform density between −2 and 2. The prior probabilities are $P(C_{+1}) = 0.6$ and $P(C_{-1}) = 0.4$.

▷ Item a

Apply Bayes' methodology to obtain the optimal classification model and show, using adiagram, the model's architecture (with clear indication of the discriminant function(s)).

▷ Item b

Calculate the model's decision boundary. Analytically calculate the classifier error probability (the integrals can be numerically solved).

▷ Item c

Recalculate the model's decision boundary if the Maximum-Likelihood criterion is adopted, this time. Analytically calculate the classifier error probability (the integrals can be numerically solved).

▷ Item d

Compare the performances of both methodologies.

A classification model is tasked with assigning a value x on the line as belonging to one of the distributions, i.e. classifying it as "+1" or "−1"; an *optimal* classification model will do so while maximizing its chances/minimizing its error.

It is given an input $x \in \mathbb{R}$ and must output a class C_B such that the conditional probability $P(C = C_B|x)$ is the greatest amongst the considered classes. Using Bayes' theorem, we can write it as

$$P(C_B|x) = \frac{p(x|C_B)P(C_B)}{p(x)} \tag{17}$$

with $p(x|C)$, called likelihood, the probability of x happening given that it comes from class C. Since the expression above is valid not only to the optimal choice but for any class, and $p(x)$ is invariant with respect to classes, we have

$$C_B = \operatorname*{argmax}_C P(C|x) = \operatorname*{argmax}_C p(x|C)P(C). \tag{18}$$

Since $x \in \mathbb{R}$, the maximization above splits the real line in compact intervals such that x in adjacent intervals result in different C_B. Classification error comes from the true class of a sample not being the class assigned to the associated interval.

For example, let's say two classes H_0 and H_1 have likelihoods such that the classifier assings $x \in R_0$ to H_0 and $x \in R_1$, to H_1. Then, the probability of error is given by the integration of $P(C = H_1|x \in R_0)$ plus the integration of $P(C = H_0|x \in R_1)$, i.e., the proportion of false negatives and positives with respect to H_1:

$$\begin{aligned} P_{\text{error}} &= P(H_1|x \in R_0) + P(H_0|x \in R_1) && (19) \\ &= P(H_1)\int_{R_0} p(x|H_1)\ dx + P(H_0)\int_{R_1} p(x|H_0)\ dx. && (20) \end{aligned}$$

Had we begun this answer by assuming a classifier maps regions R_0, R_1 to classes, the optimal classifier would minimize P_{error} by picking looking at the values of $P(H_i)p(x|H_i)$ and chossing H_i with biggest result.

Let's finally apply this Bayesian approach to the problem in question. Since there are two classes and one of them is gaussian, it is useful to rewrite the maximization problem in (18) as

$$C_B = \begin{cases} C_{+1}, & g(x) > 0 \\ C_{-1}, & g(x) < 0 \end{cases}, \quad g(x) = \log \frac{p(x|C_{+1})P(C_{+1})}{p(x|C_{-1})P(C_{-1})} \tag{21}$$

The function g, the discriminant of the classifier, is explicitly given by

$$\begin{aligned} g(x) &= \log \frac{p(x|C_{+1})P(C_{+1})}{p(x|C_{-1})P(C_{-1})} && (22) \\ &= \log[p(x|C_{+1})] + \log[P(C_{+1})] - \log[p(x|C_{-1})] - \log[P(C_{-1})] && (23) \\ &= \begin{cases} -\frac{1}{2}\log(2\pi) - \frac{1}{2}(x-2)^2 + \log(0.6) - \log(1/4) - \log(0.4), & |x| \leq 2 \\ \infty, & |x| > 2 \end{cases} && (24) \end{aligned}$$

where it was used the fact that $P(C_{+1}) = 0.6$, $P(C_{-1}) = 0.4$, $p(x|C_{+1}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-2)^2}$, and $P(x|C_{-1}) = 1/4$, if $|x| \leq 2$, and $P(x|C_{-1}) = 0$, otherwise. From (24), one can deduce that the classifier decision boundary, i.e., the roots of the discriminant, is a single point $\delta = 2 - \sqrt{\log(36/2\pi)}$; $x > \delta$ or $x < -2$ are classified as C_{+1}, and $-2 < x < \delta$ are classified as C_{-1}.

On $|x| \leq 2$, the error probability of this classifier can be calculated using (20):

$$P^B_{\text{error}} = 0.6 \int_{-2}^{\delta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-2)^2}\, dx + 0.4 \int_{\delta}^{2} \frac{1}{4}\, dx = 0.1880\ldots \tag{25}$$

As commented, this classifier minimizes the probability of error. If a different discriminant was used, such as Maximum-Likelihood (ML) criterion, error would increase. Indeed, taking

$$g(x) = g_{\text{ML}}(x) = \log \frac{p(x|C_{+1})}{p(x|C_{-1})} = \begin{cases} -\frac{1}{2}\log(2\pi) - \frac{1}{2}(x-2)^2 - \log(1/4), & |x| \leq 2 \\ \infty, & |x| > 2 \end{cases} \tag{26}$$

nudges the threshold to $\delta_{\text{ML}} = 2 - \sqrt{\log(16/2\pi)}$ and result in sligthly higher error probability:

$$P^{\text{ML}}_{\text{error}} = 0.6 \int_{-2}^{\delta_{\text{ML}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-2)^2}\, dx + 0.4 \int_{\delta_{\text{ML}}}^{2} \frac{1}{4}\, dx = 0.1967\ldots \tag{27}$$

Figure 4 shows the problem with choosing ML over the Bayesian approach: while both methods try to place a boundary between classes by looking at their likelihoods, ML ignores priors, skewing its perception of data. Had the priors been equal, however, ML would be identical to the optimal classifier. In general, the more variance between classes priors, the worse ML performs.

Figure 4: Class conditional distributions before and after being weighted by their priors.

To do!

► Question 5

Consider the dataset of 3000 bivariate, labeled samples in `data.csv` file. Firstly, split thedataset into training (70%) and test set (30%).

▷ Item a

Assume the samples labeled with "+1" are drawn by a bivariate gaussian density with parameters μ_{+1}, Σ_{+1}, while the samples labeled with "−1" are drawn by another gaussiandensity with parameters μ_{-1}, Σ_{-1}. The prior probabilities $P(C_{+1})$ and $P(C_{-1})$ are unknown as well. Estimate the missing parameters (with the training set), present the discriminantfunctions of the Bayes classifier (MAP criterion) and evaluate its performance(accuracy, precision, recall) over the test set. Comment all your results.

▷ Item b

Calculate the decision boundary and plot it on a graph with the samples of the test set. Comment your results.

► Question 6

Do your own implementation and train a multivariate linear regression model for a given problem (suggestions: Kaggle, UCI Machine Learning Repository). You may use a validation set to train models with different subsets of features and select the best one. Then, use a test set to report and comment the final model results.