Deep Learning KU (DAT.C302UF), WS25
Assignment 1
Maximum Likelihood Estimation, Decision Theory

Simon Hitzginger
simon.hitzginger@tugraz.at

Teaching Assistants:	Leon Tiefenböck
Points to achieve:	10 pts
Deadline:	05.11.2025 23:59
Hand-in procedure:	This is a **solo assignment**. No teams allowed. Submit **your report (PDF)** to the TeachCenter. You do not have to add the cover letter since there are no teams allowed.
Plagiarism:	If detected, 0 points for all parties involved. If this happens twice, we will grade the group with "Ungültig aufgrund von Täuschung"

Toy Setting – Bank

You work for a bank and are asked to create statistical models of the income of the bank's customers, and to help decide whether it is a good idea to grant them a loan.

For each customer, you are given their yearly income $x_i \in \mathbb{R}$, sampled i.i.d. from the true income distribution of the bank's customer base. The empirical distribution of observed incomes is shown in Figure 1.

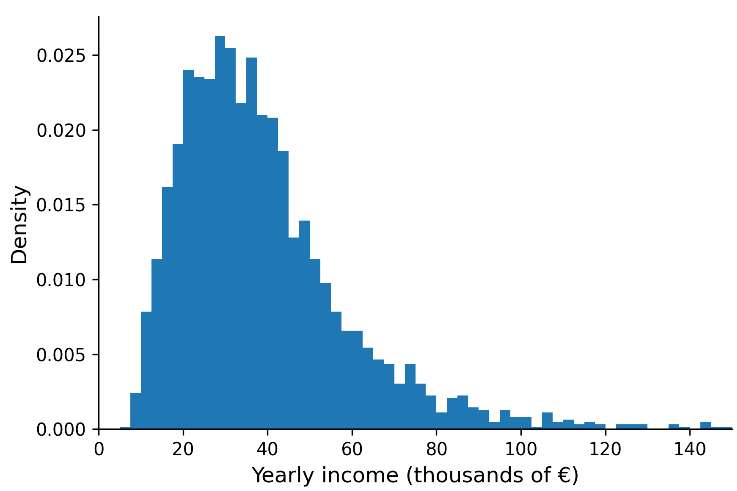

Figure 1: Histogram of observed yearly incomes.

From the shape of the data, you hypothesize that the incomes follow a *Log-Normal* distribution, whose probability density function is given by

$$p(x \mid \mu, \sigma^2) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right), \qquad x > 0. \tag{1}$$

Task 1 – Maximum Likelihood Estimation [3 Points]

The bank asks you to fit a model for the customer income distribution. Assume that the parameters μ and σ^2 are unknown.

1. Let $X = (x_1, \ldots, x_N)$ denote the observed incomes. Write down the likelihood of the entire dataset under our model $p_\theta(X)$, where $\theta = (\mu, \sigma^2)$, and express it in terms of the single-sample likelihoods $p_\theta(x_i)$. Then, take the logarithm to obtain the *negative log-likelihood* (NLL) in terms of the individual log-densities $\log p_\theta(x_i)$.
2. Using the *maximum likelihood principle*, compute the estimate $\hat{\mu}$ that minimizes $\mathrm{NLL}(\theta)$. Show that

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} \ln x_i.$$

3. Assume we have learned both parameters μ and σ^2. Briefly explain what this fitted model can be used for in practice. Give one simple example in the context of the bank scenario.

Task 2 – Conditional Maximum Likelihood [3 Points]

After modeling the income distribution in Task 1, the bank now wants to understand how income relates to the decision of granting a loan. For each applicant, you are given their yearly income x_i (as before) and a binary target $t_i \in \{0, 1\}$, where

$$t_i = \begin{cases} 1, & \text{if the applicant was granted a loan,} \\ 0, & \text{if the applicant was not granted a loan.} \end{cases}$$

The bank would like to model the probability of a loan being granted as a function of the applicant's income. We therefore define the following probabilistic model:

$$p_\theta(t = 1 \mid x) = \sigma(w_0 + w_1 \log x), \qquad \sigma(z) = \frac{1}{1 + e^{-z}}, \quad \theta = (w_0, w_1).$$

1. In Task 1, you modeled the distribution of incomes $p(x)$ using a Log-Normal model. Explain why, for predicting loan decisions, it makes sense to use the above *conditional* model $p_\theta(t \mid x)$ instead of the income model from Task 1.
2. With $X = (x_1, \ldots, x_N)$ and $t = (t_1, \ldots, t_N)$ denoting the incomes and corresponding loan-decision labels, write down the likelihood of the entire dataset under our model. Express this likelihood in terms of the single-sample likelihoods $p_\theta(t_i \mid x_i)$ under the i.i.d. assumption. Then, write down the *negative log-likelihood* and show that it is equivalent to the binary cross-entropy error

$$\mathrm{NLL}(\theta) = -\sum_{i=1}^{N} \Big[t_i \log \hat{p}_i + (1 - t_i)\log(1 - \hat{p}_i) \Big], \qquad \hat{p}_i = \sigma(w_0 + w_1 \log x_i).$$

Task 3 – Decision Theory [4 Points]

Assume we have trained a probabilistic model $p_\theta(t \mid x)$ that estimates, for each applicant $x_i \in \mathbb{R}$, the probability of a loan being granted ($t_i = 1$) or not ($t_i = 0$). We now want to make optimal decisions based on these probabilities.

1. We define the *zero-one loss function*

$$\mathcal{L}(t,\hat{t}) = \begin{cases} 1, & \text{if } t \neq \hat{t}, \\ 0, & \text{otherwise.} \end{cases}$$

where t is the true outcome (loan granted or not) and $\hat{t}$ is our decision. Write down the decision function $f : \mathbb{R} \to \{0,1\}$ that minimizes the expected loss

$$\mathbb{E}_{(x,t)}\left[\mathcal{L}(t, f(x))\right],$$

i.e., the decision rule that minimizes the expected misclassification rate.

2. Assume f additionally had access to the true marginal distribution $p^*(x)$. Could we construct a different decision function that achieves a lower expected loss under the zero-one loss? Briefly justify your answer.

3. The bank manager decides that loans should be granted more scarcely. To reflect this preference, define a new loss function $\mathcal{L}(t,\hat{t}) = L_{t,\hat{t}}$ with

$$L = \begin{bmatrix} L_{0,0} & L_{0,1} \\ L_{1,0} & L_{1,1} \end{bmatrix},$$

where $L_{t,\hat{t}}$ specifies the loss for deciding $\hat{t}$ when the true label is t.

Using this loss matrix, write down the new decision function $g : \mathbb{R} \to \{0,1\}$ that minimizes the expected loss

$$\mathbb{E}_{(x,t)}\left[\mathcal{L}(t, g(x))\right].$$

For a particular applicant x with $p(t = 1 \mid x) = 0.6$, determine the outputs of $f(x)$ and $g(x)$. Does this loss matrix achieve the goal of granting loans more scarcely? Briefly explain.