Probabilistische Methoden der Informatik

TECHNISCHE UNIVERSITÄT DARMSTADT

Wintersemester 2024 / 2025, Probe-Exam

J. Peters, T. Gruner, N. Bohlinger, L. Schulze, P. Jansonnie

Name, Vorname: ______________________ Matrikelnummer: ␣␣␣␣␣␣␣

Overview

- You have 90 minutes.
- There are 30 questions for a total of 87 points.
- Write clearly! We cannot give you points for what we cannot read.
- Solutions without derivations will not receive full points.
- Be concise! Use short sentences and get straight to the point. Unnecessarily long answers result in point deduction!
- You are allowed to use one cheat sheet with handwritten notes on both sides, but you have to submit it together with your exam (therefore, put your name on it).
- Fill in the solution on this sheet. If necessary, add sheets for the derivations and put your name and Matrikelnummer on them.

I have understood the bullet points above and agree with them. Signature: ______________________

Email: ______________________

Question 1 (2 Points)

Given a continuous random variable X with a known PDF $f(x)$ and CDF $F(x)$, explain two different methods for computing $P(a \leq X \leq b)$.

Question 2 (2 Points)

How does a linear transformation $y = ax + b$ affect the expected value and variance of a random variable?

Question 3 (4 Points)

Let a continuous random variable X follow an exponential distribution. The probability density function f_X associated to an exponential distribution is

$$f_X(x) = \lambda \exp(-\lambda x), \quad \lambda > 0, \quad x \in [0, +\infty).$$

a) Calculate the expected value of X as a function of λ. Verify that its sign makes sense.
b) Explain in one sentence why X has a non-zero skewness.

Question 4   (2 Points)

Using the Probability Axioms, prove $p(\emptyset) = 0$:

Question 5   (2 Points)

State Bayes' theorem and name each term in the equation.

Question 6 (4 Points)

A system collects data from a binary sensor and transmits it via Wi-Fi. The sensor has a 25% probability of reading 1. However, when it's raining, only 20% of the 1 readings are successfully transmitted.

a) What is the probability of receiving a 1 reading on a rainy day?
b) For redundancy, a second independent sensor is installed. What's the probability of getting a 1 as reading on a rainy day? (There are no false positive, the final reading is $S_1 \cup S_2$).
c) What is the probability that it is raining, given that both sensors report a 1 on a day with a 35% chance of rain?

Question 7 (2 Points)

Define the Shannon entropy of a discrete random variable. What is the unit of the Shannon entropy for logarithm with base 2?

Question 8 (4 Points)

Show that $D_{KL}(P\|Q) \geq 0$ and that equality holds if and only if $P = Q$.

Question 9 (5 Points)

Consider a binary erasure channel (BEC) with erasure probability ϵ (where $0 \le \epsilon \le 1$). The channel has input alphabet $\{0,1\}$ and output alphabet $\{0,1,e\}$, where e denotes an erasure. The channel transition probabilities are given by:

$$P(Y = x|X = x) = 1 - \epsilon \quad \text{for } x \in \{0,1\},$$

$$P(Y = e|X = x) = \epsilon \quad \text{for } x \in \{0,1\}.$$

Compute the channel capacity C of the BEC, showing all steps in your solution.

Question 10 (3 Points)

A call center receives customer calls at an average rate of 2 calls per hour.

a.) Suppose the random variable X models the total number of calls received in a two-hour period. What is the probability mass function of X?

b.) What is the probability that the call center receives less than 2 calls in a two-hour period? (round to two decimal places)

Question 11 (3 Points)

Let X_1, X_2 be continuous, independent random variables with probability density functions $f_{X_1}(x) = f_{X_2}(x) = e^{-x},\ x > 0$. Find the probability density function of $Y = X_1 + X_2$.

Question 12 (2 Points)

Briefly describe how the Gaussian distribution can be derived from the principles of maximum entropy and the central limit theorem.

Question 13 (2 Points)

What properties must a function $K(x)$ satisfy to be used as a kernel in kernel density estimation?

Question 14 (8 Points)

Your friend proposes a game where he rolls a die and you both have to guess the outcome. If you guess the correct number, he gives you 5 Euros. If you guess incorrectly, you give him 1 Euro.

Because your friend has tricked you before, you are suspicious and ask him to roll the die n times before you decide to play.

Let $\boldsymbol{x} \in \{1, 2, 3, 4, 5, 6\}^n$ be the sequence of observed die rolls.

From your Probability and Machine Learning (PROMI) lecture, you recall that a single roll follows a categorical distribution, where the probability of rolling a specific number k is given by:

$$P(X = k|\theta) = \theta_k, \quad \text{where} \quad \sum_{k=1}^{6} \theta_k = 1.$$

a) Show that the likelihood of observing the sequence $\boldsymbol{x}$ is given by

$$P(X = \boldsymbol{x}|\theta) = \prod_{k=1}^{6} \theta_k^{n_k},$$

where n_k is the number of times the outcome k appears in $\boldsymbol{x}$. (1 point)

b) Show that the maximum likelihood estimate (MLE) for each θ_k is given by

$$\hat{\theta}_{k,\text{MLE}} = \frac{n_k}{n}$$

. (4 points)

c) The manufacturer of the die tells you that θ follows a Dirichlet distribution:

$$P(\theta) = \text{Dirichlet}(\alpha_1, \alpha_2, \ldots, \alpha_6).$$

What is the maximum a-posteriori (MAP) estimate for θ_k? (3 points)

Hint: The Dirichlet distribution is defined as

$$\text{Dirichlet}(\theta|\alpha) = \frac{1}{B(\alpha)} \prod_{k=1}^{6} \theta_k^{\alpha_k - 1},$$

where $B(\alpha)$ is a normalization constant. In the following, the specific form of B is not needed.

Question 15 (1 Points)

What is the difference between maximum likelihood regression and Bayesian regression? Name one advantage of Bayesian regression over maximum likelihood regression.

Question 16 (2 Points)

You are working in a digital bookstore company and you would like to categorize books into predefined multiple genres. What kind of problem is that? Which approach(es) would you use for solving the problem?

Question 17 (1 Points)

In some science fiction movie, a crazy scientist has a data set with 3 data points and fits a 3rd order polynomial to it. Why would this yield a bad solution in real life?

Question 18 (2 Points)

Write down the solution of linear regression. How can we make it more robust to overfitting? Write down the modified equation and the name of the method. Name all terms.

Question 19 (4 Points)

It is well known that using the Expectation-Maximization algorithm with a Gaussian Mixture Model can lead to singularities in the estimation.
(1) Please explain why such singularities can occur and how they can be eliminated.
(2) Why does this problem not arise for a single Gaussian distribution when maximizing its likelihood function over a set of independent and identically distributed datapoints?

	Level of Significance α for One-Tailed Test								
	.25	.20	.15	.10	.05	.025	.01	.005	.0005
	Level if Significance α for Two-Tailed Test								
df	.50	.40	.30	.20	.10	.05	.02	.01	.001
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	63.662
2	.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.599
3	.765	.978	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	.741	.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	.727	.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
10	.700	.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
20	.687	.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
30	.683	.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	.681	.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
50	.679	.849	1.047	1.299	1.676	2.009	2.403	2.678	3.496
100	.677	.845	1.042	1.290	1.660	1.984	2.364	2.626	3.390
∞	.674	.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

Question 20 (5 Points)

A researcher is evaluating the performance of two sorting algorithms, `AlgA` and `AlgB`, in terms of execution time (in milliseconds). The researcher collects independent samples by running each algorithm on the same set of randomly generated input arrays. The execution times are assumed to be normally distributed, but the population standard deviation is unknown. The results from **11 trials per algorithm** are recorded as follows:

- Algorithm A: Sample mean $\bar{X}_A$, sample variance S_A^2, and sample size $n = 11$
- Algorithm B: Sample mean $\bar{X}_B$, sample variance S_B^2, and sample size $n = 11$

Assume that the population variances for both algorithms are equal.

a) State the null and alternative hypotheses for this test.

b) Write the form of the test statistic for an independent two-sample t-test. Show all formulas and steps.

c) Given a significance level of $\alpha = 0.05$, determine the critical values and describe the rejection region for a two-tailed test. Use the t-table below to compute the critical value.

d) Suppose the calculated test statistic is $t = 2.45$. Based on the rejection region, should the researcher reject the null hypothesis? Justify your answer.

Question 21 (3 Points)

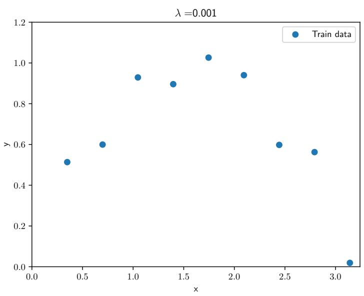
A data scientist wants to use a Z-test to compare the height of Magnolia flowers in Darmstadt to the height of Magnolia flowers in Frankfurt.

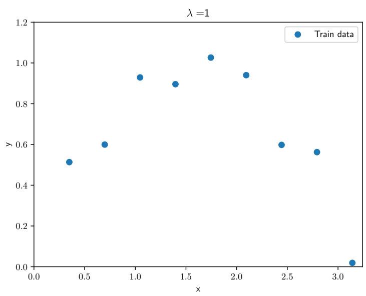
a) List and explain three key assumptions required for a Z-test to be valid.

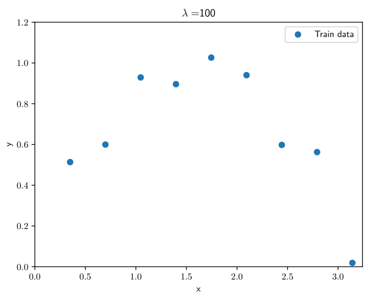
b) For each assumption, provide an example scenario where it might be violated in the context of this study.

Question 22 (3 Points)

Assume that you are provided the data below and try to estimate the underlying data generating function $f(x)$ with ridge regression based on polynomial features of degree 5. You are asked to test 3 different regularization parameters $\lambda = \{0.001, 1, 100\}$. Please draw the approximation function that you would get for the three different regularization parameters in the provided plots. Note that we don't expect you to draw accurate plots, but want to see if you have grasped the concept of ridge regression. Please shortly explain the phenomena that you would expect for each of the regularization parameters.







Question 23 (1 Points)

Is a maximum likelihood estimator always unbiased? Please explain your answer or give a counterexample.

Question 24 (2 Points)

Suppose that your estimator has a good training performance but you observe poor test performance. You decide to use a highly flexible model capable of approximating any function. Why might this be problematic with a finite training dataset, and how can you mitigate this issue?

Question 25 (2 Points)

Let $\{x_n\}_{n=1}^N$ be a sequence of random variables from an unknown distribution $p(x_1, \ldots, x_N)$. Given that the first moment $\mathbb{E}[x_n]$ exists, but the sample mean

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$$

does not converge, explain a possible reason for this based on the conditions of the Law of Large Numbers.

Question 26 (2 Points)

Both pseudorandom number generators (PRNGs) and quasi-random number generators (QRNGs) produce random number sequences. Explain how their behavior differs in terms of randomness and uniformity.

Question 27 (2 Points)

Direct sampling is a Monte Carlo method used for estimating point estimates from probability distributions. Under what conditions is direct sampling applicable, and why is it not always feasible?

- Applicable when the inverse cumulative distribution function (CDF) can be computed analytically, allowing direct sampling:

  $$x = F^{-1}(u), \quad u \sim U(0,1)$$

  (1.0)
- Not always feasible because many complex distributions lack a closed-form inverse CDF, making numerical computation computationally expensive or even impractical. (1.0)

Question 28 (2 Points)

Explain the difference between the Q-function and the value function in Reinforcement Learning.

Question 29 (3 Points)

A Markov Decision Process is defined by five terms:

- S: the state space.
- A: the action space.
- P: the transition kernel.
- R: the reward function.
- γ: the discount factor.

Which ones of them can be used to model the uncertainties of a system? Justify your answer.

Question 30 (7 Points)

You are a robotics fleet engineer working on a team developing autonomous robotic arms for an assembly line. These robotic arms need to pick up objects and place them with high precision. Your team has designed several grasping strategies. Some strategies are well-tested and work reliably across most objects. Other strategies are experimental and could potentially improve efficiency and reduce errors, but their performance is uncertain.

a.) Discuss the decision of choosing a specific grasping strategy based on the exploration-exploitation trade-off. (2.0 point)

b.) Your manager asks you to clarify the types of uncertainty involved. Explain them and give one example each. (1.0 point)

c.) Your manager is unsure if he understood what you were explaining and asks about the consequences of not handling uncertainty in this context. Name two consequences, each with one example. (1.0 points)

d.) Your manager got nervous and asked you to take immediate measures. However, he reminds you that testing new grasping strategies takes a lot of time and money. To tackle this, you decide to use Bayesian decision-making. Explain what kind of surrogate model to use, and sketch the algorithm. (2 points)

e.) Your manager asks why you opted for a surrogate model instead of an approach without one. Explain the advantages of using a surrogate model in this context and the limitations of approaches that do not rely on one. (1 point)