

Name:					
Matriculation Number:					
Tutorial Group:	A <input type="checkbox"/>	B <input type="checkbox"/>	C <input type="checkbox"/>	D <input type="checkbox"/>	E <input type="checkbox"/>

Question:	1 (5 Points)	2 (5 Points)	3 (5 Points)	4 (5 Points)	Total (20 points)
Score:					

General instructions:

- The written test contains 4 questions and is scheduled for 45 minutes. The maximum amount of points you can earn is 20.
- Please verify if your exam consists of 12 pages with 4 questions printed legibly, else contact the examiner immediately.
- No electronic devices (calculator, notebook, tablet, PDA, cell phone) are allowed.
- Answers without sufficient details are void (e.g.: you can't just say "yes" or "no" as the answer).
- Last page consists of material that you may use to solve the questions. You may detach the last page for your convenience.
- You will be provided additional working sheets if necessary. Make sure to return them along with your solution sheet.
- Please provide your ID card when asked by the examiner.
- Please fill in name, matriculation number (student registration number) and tutor group in the form above and return the solution sheets into the provided box.
- Please sign below.

Student's Signature _____

D5: DATABASES AND INFORMATION SYSTEMS
INFORMATION RETRIEVAL AND DATA MINING, WS 2013/14
DR. KLAUS BERBERICH AND DR. PAULI MIETTINEN
FIRST SHORT TEST, DURATION: 45 MINUTES



LINEAR ALGEBRA

Problem 1. Consider the following matrix A ,

$$A = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}.$$

a) Consider the first and second column of A ,

$$\begin{aligned} \vec{a}_1 &= (1/\sqrt{2}, 1/\sqrt{2})^T \\ \vec{a}_2 &= (1/\sqrt{2}, -1/\sqrt{2})^T. \end{aligned}$$

Compute the Euclidean (that is, L_2) norm of these two vectors. [**2 points**]

b) Compute the dot product $\vec{a}_1 \cdot \vec{a}_2$. [**1 point**]

c) Is A invertible? If yes, give its inverse. If not, explain why. [**2 points**]

d) Let I be the 2-by-2 identity matrix and let

$$\vec{v} = (1, 0)^T.$$

Compute

$$AIAI(A^T)^2IAIIAIA^T\vec{v}.$$

[**1 point**]

Solution

D5: DATABASES AND INFORMATION SYSTEMS
INFORMATION RETRIEVAL AND DATA MINING, WS 2013/14
DR. KLAUS BERBERICH AND DR. PAULI MIETTINEN
FIRST SHORT TEST, DURATION: 45 MINUTES



Problem 2. Let A be a 500-by-100 matrix and let $A = U\Sigma V^T$ be its SVD.

- a) What are the sizes (numbers of rows and columns) of U , Σ , and V ? [1 point]
- b) How many non-zeros can matrix Σ have at most? [1 point]
- c) Is matrix A invertible? Explain why/why not. [1 point]
- d) Assume that $\text{rank}(A) = 100$ and consider the matrix $\tilde{A}^+ = (A^T A)^{-1} A^T$ (you can assume that the inverse exists). Prove that \tilde{A}^+ is the pseudo-inverse of A . You can use the fact that $V\Sigma^+U^T$ is the pseudo-inverse of A , that $(XY)^T = Y^T X^T$ for all matrices X and Y for which the product is well-defined, and that $(XY)^{-1} = Y^{-1} X^{-1}$ if X and Y are invertible. [2 points]

Solution

PROBABILITY THEORY

Problem 3. Let X and Y be two discrete random variables such that X takes values from $\{1, 2, 3, 4\}$ and Y takes values from $\{3, 6, 12\}$. Let their joint mass function $f_{X,Y}$ be as follows:

		X			
		1	2	3	4
Y	3	1/6	0	1/12	1/12
	6	1/6	1/6	1/12	1/12
	12	0	0	1/12	1/12

- What is the marginal distribution of Y ? [**1 point**]
- What is the expected value of Y , $E[Y]$? [**1 point**]
- What is the conditional expectation of Y given X , $E[Y | X]$? [**2 points**]
- Let A be a random variable with $E[A] = 4$ and let B be a random variable with $E[B] = 6$. Let C be a random variable defined as $2(A + B) - 20$. What is $E[C]$? [**1 point**]

Solution

D5: DATABASES AND INFORMATION SYSTEMS
INFORMATION RETRIEVAL AND DATA MINING, WS 2013/14
DR. KLAUS BERBERICH AND DR. PAULI MIETTINEN
FIRST SHORT TEST, DURATION: 45 MINUTES



STATISTICAL INFERENCE

Problem 4. Suppose we have the following sample of 20 response times from a search engine

$$X = \langle 10, 9, 1, 8, 2, 7, 3, 6, 4, 5, 5, 2, 3, 4, 1, 5, 8, 2, 10, 5, 7, 3, 4, 6, 5 \rangle$$

- (a) What are the sample mean \bar{X} and the sample variance S^2 ? [1 point]
- (b) What is the 95% confidence interval of \bar{X} (assuming $\sigma^2 = 1.44$)? [2 points]
- (c) Is there *strong evidence* to reject the null hypothesis $H_0 : \mu = 5.5$ (assuming $\hat{s}e = 0.25$)? [2 points]

Solution

ADDITIONAL MATERIAL

Linear algebra

- Identity matrix: n -by- n matrix I such that $I_{ij} = 1$ iff $i = j$ and $I_{ij} = 0$ otherwise
- Product with identity matrix: $AI = IA = A$ for all n -by- n matrices A
- Matrix inverse: $A^{-1}A = AA^{-1} = I$
- Transpose identities: $(A^T)^T = A$ for all A ; $(AB)^T = B^T A^T$ when the product is well-defined
- Inverse of a product: $(AB)^{-1} = B^{-1}A^{-1}$ if A and B are invertible
- Inverse of orthogonal matrices: $A^T = A^{-1}$ iff A is orthogonal

Probability & Statistics:

- Bayes' Theorem: $\Pr[A|B] = \frac{\Pr[B|A] \Pr[A]}{\Pr[B]}$
- Law of Total Probability: $\Pr[B] = \sum_{i=1}^n \Pr[B|A_i] \Pr[A_i]$ for disjoint events A_i with $\sum_{i=1}^n \Pr[A_i] = 1$
- Expectation: $\mathbf{E}[X] = \sum_{k=1}^n k f_X(k)$ and Variance: $\mathbf{Var}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2$ for a discrete RV X with density function f_X
- Markov inequality: $\Pr[X \geq t] \leq \frac{\mathbf{E}[X]}{t}$ for $t \geq 0$ and a non-neg. RV X
- Chebyshev inequality: $\Pr[|X - \mathbf{E}[X]| \geq t] \leq \frac{\mathbf{Var}[X]}{t^2}$ for $t > 0$ and a non-neg. RV X
- Sample Mean: $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ and Sample Variance: $S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$
- For an estimator $\hat{\theta}$ of parameter θ over i.i.d. samples $\{X_1, X_2, \dots, X_i, \dots, X_n\}$,
 - If $\mathbf{E}[X_i] = \mu$, then $\mathbf{E}[\hat{\theta}_n] = \mu$
 - If $\mathbf{Var}[X_i] = \sigma^2$, then $\mathbf{Var}[\hat{\theta}_n] = \frac{\sigma^2}{n}$
 - Standard Error: $se(\hat{\theta}) = \sqrt{\mathbf{Var}[\hat{\theta}_n]}$
 - Mean Squared Error: $MSE[\hat{\theta}_n] = (\mathbf{E}[\hat{\theta}_n] - \theta)^2 + \mathbf{Var}[\hat{\theta}_n]$

