CISPA HELMHOLTZ CENTER FOR INFORMATION SECURITY

UNIVERSITÄT DES SAARLANDES

Problem 1 (Probability Theory).

1. Calculate the sample covariance between X and Y in the following data (one observation per row).

X	Y
9	1
3	-9
1	-1
4	3
10	-9
5	-1
9	-3
3	-10
10	-4
3	-2

2. What would be the new value for the sample covariance if:

   - Column X was multiplied by 0.5?
   - Column Y was replaced by a value $2c - 0.5$, where c is a constant that is fixed for all Y.

Solution.

1. Covariance: -0.55

2.
   - $Cov(aX, Y) = a\ Cov(X, Y) \longrightarrow 0.5 \cdot (-0.55) \longrightarrow -0.275$
   - If c is exactly the same for all the values of Y, Y is a constant. Denoting $2c - 0.5$ by k we have: $Cov(X, Y) = Cov(X, k) = 0$ since covariance of a random variable and a constant is zero.

Problem 2 (Probabilities). What are the canonical parameters associated with the following probability distributions?

- Uniform
- Bernoulli
- Binomial
- Exponential
- Gaussian

Solution.

- Uniform: Lower limit a and upper limit b
- Bernoulli: Probability of success p
- Exponential: Rate variable λ
- Binomial: Number of tries n, Probability of success p
- Gaussian: Mean μ and standard deviation σ

Problem 3 (Probabilities). Let X_1 and X_2 be two independent binary random variables with $\Pr(X_1 = 1) = 0.8$ and $\Pr(X_2 = 1) = 0.5$. Let Y be their sum, $Y = X_1 + X_2$.

1. What is $\mathbb{E}(X_1)$?

2. What is $\mathbb{E}(Y)$?

3. What is $\mathbb{E}(Y \mid X_2 = 0)$?

Solution.

1. $\mathbb{E}(X_1) = 0.8$

2. $\mathbb{E}(Y) = \mathbb{E}(X_1 + X_2) = \mathbb{E}(X_1) + \mathbb{E}(X_2) = 0.8 + 0.5 = 1.3$
   N.B.[1] the linearity of expectation.

3.

$$\mathbb{E}(Y \mid X_2 = 0) = \sum_{y=0}^{2} y \cdot \Pr(Y = y \mid X_2 = 0) = 0 \cdot \Pr(Y = 0 \mid X_2 = 0)$$
$$+ 1 \cdot \Pr(Y = 1 \mid X_2 = 0) + 2 \cdot \Pr(Y = 2 \mid X_2 = 0)$$
$$= \Pr(Y = 1 \mid X_2 = 0) = \Pr(X_1 = 1) = 0.8$$

Problem 4 (Analysis).   What do the following functions converge to?

1. $f(x) = x/(1 - x)$ for $x \to \infty$

2. $f(x) = \frac{x^2 - 4}{x - 2}$ for $x \to 2$

3. $f(x) = (n + 1)x/x^n$ for $n > 1$ and for $x \to \infty$

4. $f(k) = \sum_{i=1}^{k}(1/2)^k$ for $k \to \infty$

Solution.

1. -1
   Make use of rule of L' Hospital.

2. 4
   Factor the nominator, then cancel the denominator.

3. 0
   Again, de L' Hospital's rule.

4. 1
   We know that the geometric series (starting from $i = 0$) is equal to $\frac{1}{1 - \frac{1}{2}} = 2$. The result follows
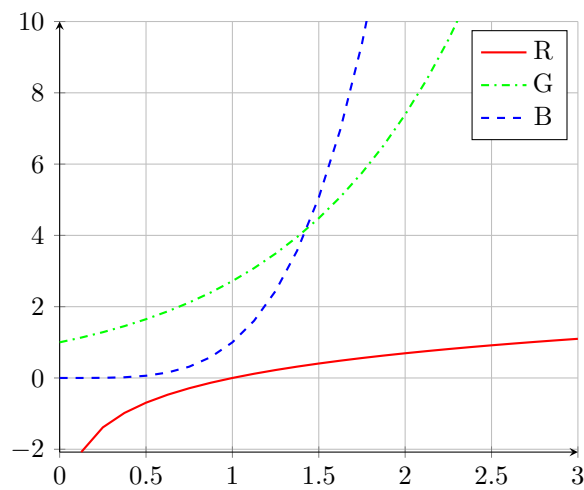   since this series starts with index $i = 1$.

Problem 5 (Analysis).
Mark which of the following three functions correspond to which of the lines in the plot below.

---

[1]NB is the abbreviation of *nota bene*, which in Modern English means note well or pay special attention to.

1. $f(x) = e^x$

2. $f(x) = x^4$

3. $f(x) = \ln(x)$

Solution.

1. green   2. blue   3. red

Problem 6 (Combinatronics).

1. How many different subsets does a set of size n have?

2. How many different subsets of size k does a set of size n have?

Solution.

1. 2^n.   2. $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

Problem 7 (Linear Algebra).
How many solutions are there for each of the given system equations? Why?

a) $2x + y = 1$
$\quad 5x - 2y = 3$

b) $3u - 5v + 2w = 10$
$\quad 2u - 3v + 7w = 23$

c) $3p + 5q - r = 1$
$\quad -p + 4q + 2r = 9$
$\quad 6p + 10q - 2r = 3$

Solution.

1. One unique solution   2. Infinitely many solutions   3. No solutions, inconsistent system

CISPA
HELMHOLTZ CENTER FOR
INFORMATION SECURITY

UNIVERSITÄT
DES
SAARLANDES

Problem 8 (Linear Algebra).   Consider the following vectors.

1. What are the dimensions of each vector?

(a) $x_a = \begin{pmatrix} 4 \\ 2 \\ -5 \end{pmatrix}$

(b) $x_b = \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix}$

(c) $x_c = \begin{pmatrix} 5 \\ 0 \end{pmatrix}$

(d) $x_d = \begin{pmatrix} 1 \\ 1 \\ 4 \\ 2 \end{pmatrix}$

(e) $x_e = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$

(f) $x_f = \begin{pmatrix} -1 \\ 1 \\ 1 \\ -2 \end{pmatrix}$

2. Perform the following vector operations using the vectors specified in Part. 1.

(a) $x_a + x_c$

(b) $4 \cdot x_e$

(c) $x_d \cdot x_f$

(d) $\|x_d\|$

3. Which pair of vectors in Part. 1 are orthogonal?

Solution.

1. (a) 3
   (b) 3

   (c) 2
   (d) 4

   (e) 2
   (f) 4

2. (a) $\begin{pmatrix} 7 \\ 1 \\ -4 \end{pmatrix}$

   (b) $\begin{pmatrix} 0 \\ 8 \end{pmatrix}$

   (c) 0

   (d) $\sqrt{22} \approx 4.7$

3. Pairs (x_a, x_c) and (x_d, x_f).

Problem 9 (Programming).
The purpose of this exercise is to familiarise yourself with the programming environment used for this course. For the needs of this exercise download and unpack the accompanying zipped data here.

1. Ensure that you have a working *python* environment including an option for editing and running *jupyter* notebooks. For this you may consult the instructions in the course site.

2. Familiarise yourselves with the python language, to be able to use

   - native data structures
   - function and class declarations
   - execution flow control (for, if, while), and also
   - comprehensions (for generators, lists, and dicts).

   In this task you might find helpful the introductory notebook `a0_notebook9.2` that we provide; open it in your environment and study its sample source code.
   If you are already familiar with these, you may skip this question.

3. Familiarise yourselves with ML-related operations, until you are able to use

   - numpy arrays creation

- numpy array indexing
- vectorised operations on numpy arrays (sum, mean, comparisons, etc), as well as
- matrix operations (dot, matrix multiplication).

Proceed with basic plotting using the standard matplotlib package, including

- line plots
- scatter plots, and
- bar plots.

In this task you might find helpful the introductory notebook `a0_notebook9.3` that we provide; open it in your environment and study its sample source code.
If you are already familiar with these, you may skip this question.

4. Proceed with tabular data operations using the pandas package. For this you may follow the official 10 minute tutorial here.

5. Perform some basic data analysis.
   Open notebook `a0_notebook9.5` in your environment and complete the source code in the indicated cells.
   Note: You do not need to submit the resulting notebook for this self-assessment assignment; in the future, however, you will be asked to provide your completed notebook, including a printout of it as a pdf file.