T-61.5010 Information Visualization Examination

March 4th, 2013

To pass the course you must also pass the term project (assignment). Results of this examination are valid for one year after the examination date. To get full points you must complete all of the problems 1-5. All of the problems have an equal weight. There are 50 points in this exam in total.

Answer in English. Please write clearly and leave a wide left or right margin. No extra material (calculator, lecture notes etc.) is allowed. Please write your answers preferably using a ballpoint pen, not a pencil.

PLEASE ANSWER EACH QUESTION ON A DIFFERENT SHEET!

Instructions for the essays: Write in full sentences and structure your answer in paragraphs. The essay should be written in a manner understandable to your fellow student (who would have the necessary pre-requisite information to take this course, but has not taken it) who has asked you to tell him/her about the topic of the essay.

The results will be posted to the course Noppa home page. No other announcement will be made.

There are 3 (three) pages in this examination. You can keep this paper.

1 Multiple choice questions

The following questions each have different proposed answers. Only one of them is correct. **You have to give your answer along with your confidence ("High" or "Low") for each answer.** Grading for each of these multiple choice questions is then:

- +2 if answer is right and confidence is high
- +1 if answer is right and confidence is low
- 0 if answer is missing
- -1 if answer is wrong and confidence is low
- -2 if answer is wrong and confidence is high

Write on your answer sheet the correct answer (A, B, C, D, ...), along with the confidence you have (High, Low) for that questions; e.g. "A, Low" is a proper way of answering a question. **Missing confidence for a question will be treated as "Low". Total score for this question is between 0 and 10 (TOTAL SCORE OF THE EXAM IS UP TO 50).**

Question 1.

During the training of the Self Organizing Map (SOM), in each iteration a winning node is selected and prototypes are then updated. Consider a node u whose prototype vector in the data space is $\mathbf{u}$. Which of the following affirmations is NOT TRUE about the SOM update rule for $\mathbf{u}$?

A) The update rule involves a function of the iteration t.
B) The update rule involves a function of the distance between u and the winning node v on the SOM grid.
C) The update rule involves a function of the distance between $\mathbf{u}$ and the prototype $\mathbf{v}$ of the winning node v in the data space.
D) The update rule involves a function of the difference between $\mathbf{u}$ and the selected data point $\mathbf{x}$.

Question 2.

According to the CIE system of color standards,

A) In the CIE chromaticity diagram, any set of three colored lights specifies a triangle. All points within the triangle can be represented as a mixture of the given lights.
B) In the CIE chromaticity diagram, none of the realizable colors fall within the spectrum locus (the set of chromaticity coordinates representing single wavelength colors).
C) The CIE XYZ tristimulus color space is perceptually uniform.
D) All the answers above are correct
E) None of the above answers is correct

Question 3.

We want to perform Laplacian Eigenmap (LE) to reduce dimensionality of a data set. We have the graph laplacian $\mathbf{L}$ for the data set:

$$\mathbf{L} = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

We then compute the eigenvectors in the form of the matrix $\mathbf{V}$ and the eigenvalues in the form of the matrix $\mathbf{D}$, for example in Octave/Matlab, as $[\mathbf{V}, \mathbf{D}] = eig(\mathbf{L})$. We obtain (rounded values)

$$\mathbf{D} = \begin{pmatrix} -3.7 & 0 & 0 & 0 & 0 & 0 \\ 0 & -3.0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -2.0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1.0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \text{and } \mathbf{V} = \begin{pmatrix} 0.1 & 0.3 & 0.4 & -0.5 & 0.6 & 0.4 \\ -0.4 & -0.6 & -0.4 & 0.0 & 0.4 & 0.4 \\ 0.6 & 0.3 & -0.4 & 0.5 & 0.1 & 0.4 \\ -0.6 & 0.3 & 0.4 & 0.5 & -0.1 & 0.4 \\ 0.4 & -0.6 & 0.4 & 0.0 & -0.4 & 0.4 \\ -0.1 & 0.3 & -0.4 & -0.5 & -0.6 & 0.4 \end{pmatrix}.$$

Which of the following affirmations is correct:

A) Laplacian Eigenmap does not work in this case
B) Projecting to three dimensions using LE is necessary to preserve the proximity coded in $\mathbf{L}$
C) Projecting to two dimensions using LE is necessary to preserve the proximity coded in $\mathbf{L}$
D) Projecting to one dimension using LE is enough to preserve the proximity coded in $\mathbf{L}$
E) None of the above answers is correct

Question 4.

In human perception, arbitrary symbols are

A) Culture-specific and capable of rapid change and resistant to instructional bias
B) Culture-specific and capable of rapid change and hard to learn
C) Cross-cultural and capable of rapid change and hard to learn
D) Processed quickly and understandable without learning
E) None of the above answers is correct

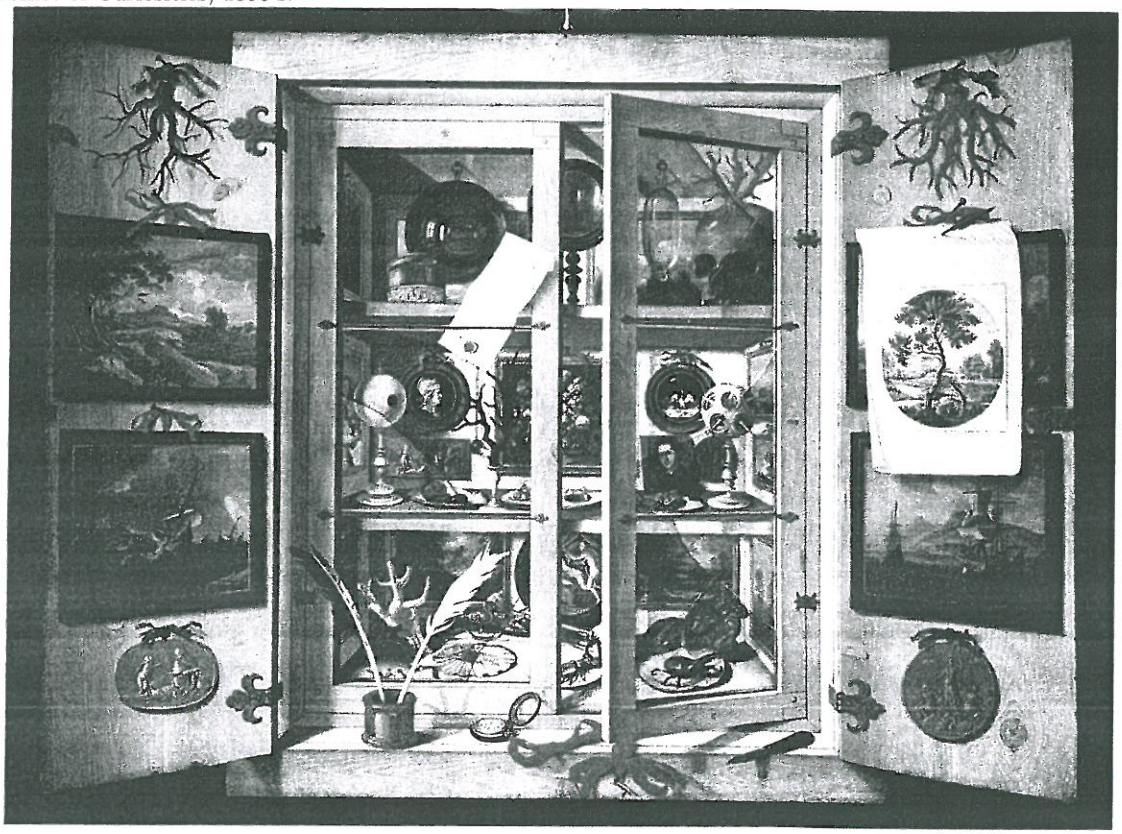
Question 5.

In the theory of pre-attentive features, the conjunction search of two pre-attentive attributes

A) is always pre-attentive
B) is sometimes pre-attentive
C) is never pre-attentive
D) does not exist
E) None of the above answers is correct

2 Depth Cue Theory

Discuss and explain the Depth Cue theory, on the basis of the piece of art below - Domenico Remps, Cabinet of Curiosities, 1690's.



3 Concept definitions

Define and explain in a maximum of 3 lines (3 lines per concept) all of the following concepts:

- Gestalt laws
- Data-ink ratio
- Pre-attentive processing
- Absolute Multidimensional Scaling
- Chart Junk

4 Essay 1

Let us assume that you have a high dimensional data set that you want to analyze. You want to perform dimensionality reduction and visualization of the reduced data. Describe, in detail, the methodology you will follow and the reasons for your choices. Include the pre-processing steps, post-processing steps, quality criteria, dimensionality reduction methods, etc. that you would use. Maximum 2 pages, explain all the technical terms that are used in your essay.

5 Essay 2

Present the Theory of Data Graphics by Edward R. Tufte and discuss its five principles.

NOTE: Both essays are compulsory!