CS306-Data analysis and visualization
Mini-project, Nov. 2016

Learning outcome:

To work on a reasonably challenging and practically relevant problem, and apply few concepts learnt during the course.

Problem description:

Video quality measurement is important in real-time applications such as video broadcasting and widely used by video content providers and distributors (eg. Netflix). The underlying idea is to achieve a more holistic real-time monitoring of video transmission based on user dissatisfaction levels. Obviously, higher user dissatisfaction will directly affect business not only in terms of existing customer base but also negatively impact future enrolments. While mean opinion score (MOS) is widely used to measure (subjectively or using an objective model) video quality, there are few limitations to it. Particularly, MOS being an averaged value may not be able to effectively quantify actual user dissatisfaction levels. As a result, new measures to supplement MOS should be used. One such simple measure is the *PDU* (percentage dissatisfied users) and is defined as

$$PDU = \frac{\#(OS < th)}{N} \times 100$$

where *OS* denotes the opinion score from an individual observer, *th* is a user-defined threshold, and *N* is the total number of observers evaluating the given condition (service quality). Thus, *PDU* simply quantifies the percentage of users who are unhappy about video quality and it can be controlled based on the threshold *th* which can be customized by the service provider. *PDU* is a useful quantity because a dissatisfied/unhappy customer/user is potentially detrimental to the service provider (broadcaster) for 2 reasons. First, he/she is likely to discontinue the service (loss of revenue), and might enroll with a competitor. Second, the negative publicity may affect prospective customers.

Tasks:

In this *mini project*, our aim is to analyze the relationship between MOS and PDU, and in the process develop prediction models. You are provided data_CS306.mat. ‘data_144’ contains subjective evaluations of 144 HDR videos (this is the same data as lab9), ‘data_100’ contains subjective evaluations of 100 full HD videos encoded using H.264/AVC. Note that two datasets are not related to each other in any sense (videos evaluated, observers, time of experiment etc. are all different) except that the same rating scale was used in both of them (ACR, 1-worst, 5-best).

1. Our first goal is to analyze the relationship between MOS and PDU. For that purpose, use 2D plots (scatter plot) and the linear correlation coefficient, to comment on the relationship between the two (MOS and PDU). You should use both the datasets (‘data_144’ and’ ‘data_100’) to arrive at your conclusions. For your plots, let MOS be plotted along x-axis.

2. In this part, we aim to obtain a prediction model for PDU given a MOS value (i.e. we need to predict the PDU from the MOS). To that end, consider three different models:

a) Linear model i.e. $PDU_{predicted} = a \times MOS + b$ where a and b are parameters to be determined via minimizing the squared error between predicted and actual PDU. (hint: use *regress* in Matlab)
b) Logistic model $PDU_{predicted} = \alpha_1 . \left(\frac{1}{2} - \frac{1}{1 + e^{\alpha_2 . (MOS - \alpha_3)}} \right) + \alpha_4$ where α_1 to α_4 the parameters of the 4-parameter logistic model, and computed by minimizing the squared error between predicted and actual PDU. (hint: use *fminsearch* in Matlab).
c) Gaussian model (hint: notice how the definition of PDU is similar to the area under the curve which is given by $P(X < x)$).

3. Perform an F-test to determine which model of the previous part (linear, Gaussian and logistic) is statistically better for analysis of MOS-PDU relationship.

4. Compare the three models (linear, Gaussian and logistic) using nonparametric testing (eg. bootstrapping), and find which is statistically better.

5. Now develop/train the above 3 models using 'data_100', and use the resultant models to predict the PDU values in 'data_144'. Using the mean squared error between the predicted and actual PDU as the criterion, which model (linear, Gaussian or logistic) is best one for predicting PDU from MOS?

Note: for parts 3 and 4, you will need to do three separate comparisons i.e. linear with logistic, linear with Gaussian, and Gaussian with logistic.

Practicalities:

- You will work groups of 2 or 3 (same as your lab).
- You will be evaluated based on few questions (based on your solutions to the tasks above and also relevant topics covered during the course) during presentation.
- Your presentation should have at most 6 slides (1 for problem definition and 1 slide for each of the 5 tasks). Expect interjection during the presentation. Time limit per group: ideally 12 mins (including Q & A) but will be increased if required for further questions.
- Members of the same group can receive different marks depending on the performance during presentation.
- Main criteria for evaluation will be the level of your understanding of the problem and the solution devised. So while certain minimum work is expected from each group, there will be incentive for more innovative ideas/solutions.
- Suggestion: in your presentation try (as much as possible) to provide visual examples (eg. plots, figures, images, comparative diagrams etc.) to explain your point. Avoid too many words/sentences on one slide.
- You will need to submit your slides a day before the presentation.
- As with any real data analysis exercise, making assumptions (eg. if you think any information is missing) is an acceptable part of the game. You will, however, be asked to justify those during your presentation.