Pattern Classification and Recognition
ECE 681

Spring 2016
Course Project 50% Status Report Guidance

Due: 4:30PM Thursday, April 7[1]

Your status report is not considered submitted until both components (hard-copy and `.pdf` file) have been submitted.

1. Submit a **hard-copy** to the homework box in Teer (for us to mark-up and return to you).
   - Staple your report (*i.e.*, do not submit loose pages, or fold and tear the corner).
     Staplers are available in Hudson 130. We cannot be responsible for keeping track of loose pages.
2. Submit a **.pdf file** as an Attachment to the Assignment in Sakai (for us to refer to when evaluating future project report submissions).

Project 50% status reports will be scored according to the following criteria, much of which is described in the course project information. As described in the course project information, the 50% status report is expected to document significant steps toward addressing the problem you have selected, including preliminary classifier performance results.

The 50% status report should be viewed as a draft of your final report, and sections that should be in final form will be scored according to the criteria for the final report. (Feedback at this stage will allow you to improve these sections, to improve your score on your final report.) It is acceptable to re-use text you originally wrote for your project proposal or 25% status report in your 50% status report, though it is more likely you will revise and expand upon the text you have already written.

There are no page requirements or limits – the 50% status report should be parsimonious (as long as it needs to be to fully describe what you have done and plan to do, but no longer than necessary). In particular, do not describe methods which you are not utilizing as part of your project.

The relative weighting of each section in the final report will be the same as the relative weightings given here for the 50% status report.

[1] I (Dr. Tantum) will collect status reports from the homework box after 4:30PM, allowing for an appropriate grace period.
DO NOT submit late status reports to the locked homework box in Teer.
Late status reports are to be submitted to me (Dr. Tantum) in person, to my personal mailbox in Hudson 130 (not the mailbox labeled ECE 681), or slid under my office door (Hudson 114).

Submitted by 4:30PM, Friday, April 8, 2016 = 1 day late. (10 points deducted from raw score)
Submitted by 4:30PM, Monday, April 11, 2016 = 2 days late. (20 points deducted from raw score)
Submitted by 4:30PM, Tuesday, April 12, 2016 = 3 days late. (30 points deducted from raw score)
Late submissions not accepted after 4:30PM, Tuesday, April 12, 2016.

Status reports submitted in person to me (Dr. Tantum), to my personal mailbox in Hudson 130 (not the mailbox labeled ECE 681), or slid under my office door (Hudson 114) after the submission deadline but prior to my collecting status reports from the homework box will be treated as if they were submitted on time.
Status reports submitted to the homework box after I have collected status reports from the box will receive zero credit.

15% Presentation Clarity

Presentation clarity will be scored according to the expectations for the final report.

Present a status report that is clearly written, easy to follow, and complete as a stand-alone document, as would a report delivered to a customer who hired you to complete this project. Someone who is not taking (or has not taken) this class, but is familiar with probability/statistics and linear algebra, should be able to read your status report and understand conceptually what you have done.

This is a formal report. As such, writing (organization, sentence structure, grammar, etc.) matters, and the presentation clarity score will reflect the quality of the writing. Make use of the writing studio to help you improve your presentation.
`http://twp.duke.edu/twp-writing-studio`

Specific pieces of advice:

1. Do not list and answer each of the "things to think about" questions.
2. Define acronyms the first time they are used.
3. Make use of section (and subsection) headings.

10% Problem Description

The problem description will be scored according to the expectations for the final report.

The problem description should be complete and thorough. Someone reading your report should have a good understanding of the problem you are considering solely from your description of it, and not need to refer to outside references in order to understand it, though references to relevant outside sources should be provided.

Provide background and context for the problem you have selected, and motivation for considering it. The motivation for considering a problem should be more significant than, "it is a classic pattern classification data set." A well-written problem description is likely to span two-thirds of a page, or more.

Some questions to think about are:

- Why is this an interesting or important problem?
- What is its significance?
- Why might others be interested in this problem?
- How might others be impacted by the results?
- What have others done to address this problem?
- What are the broader implications of successfully addressing this problem?

10% Data Description

The data description will be scored according to the expectations for the final report.

The data description should be complete and thorough. Someone reading your report should have a good understanding of the data you are using solely from your description of it. Visualizations are key; there is a reason behind the saying, "a picture is worth a thousand words."

Provide a description of the data you are considering, its source (including reference(s)), how it was collected or generated, and why it is a good choice for the problem you have selected.

It is expected that you will present both summary statistics (number of data points, proportion of data in each class, etc.) and visualizations of the data. Thoughtfully consider how you present the summary

statistics; depending on the nature of the summary statistics, bar charts may be more appropriate than tables of numbers. Visualizations of the data may include scatter plots of the features (if the data source is feature-based) or visualizations of the raw data from which features are extracted/generated.

Some questions to think about are:

- What data are you using, and what is its source?
- Why are you using this data?
- How does this data support addressing the problem you have selected?
- What was the data collection/generation procedure?
  (Whether collected/generated by you or someone else.)
- How do visualizations of the data inform your decisions regarding pre-processing, or choice of classifier(s)?

20% Approach/Methods

Pre-Processing (10%)

The pre-processing description will be scored according to the expectations for the final report.

The pre-processing description should be complete and thorough. Someone reading your report should have a good understanding of the pre-processing you did, including the mathematics behind the pre-processing and the benefits it conveyed, without referring to outside sources, though references to relevant outside sources should be provided.

Provide a description of and motivation for your pre-processing, including data normalization, how you generate features from the raw data, and/or how you perform feature selection.

It is expected that key equations will be included in the pre-processing description. For example, it is not enough to say ZmUv normalization was applied. Include the equation(s) to show how normalized features were obtained from the original (unnormalized) features.

Visualizations are key here, as well, especially comparisons of the data (features) before and after pre-processing.

Some questions to think about are:

- What have you done?
- Why were these things helpful or necessary?
- How did these things to improve the classification system, or benefit your study?

Classifiers (10%)

The classifier description(s) will be scored according to the expectations for the final report.

The classifier description(s) is/are expected to be complete and thorough, and include the mathematics of the classifier(s) you use. Someone reading your report should have a good understanding of the classifiers you used, including the mathematics behind them, without referring to outside sources, though references to relevant outside sources should be provided. Visualizations are key here, as well, and visualizations of decision surfaces are especially helpful.

Provide a description of how you approached developing classifier(s), including any design trade-off decisions you made. Be clear about what classifier(s) you coded yourself, and for what classifier(s) you used

(or leveraged) a package written by someone else (*i.e.*, a "black box"). We want to make sure you are doing more in your project than providing inputs to and scoring outputs from someone else's "black box".

Some questions to think about are:

- What characteristics did you look for in a classifier, or classifiers?
- What did you consider when choosing classifier parameters?
- What did you consider when training the classifier(s) and evaluating their performance?

15% Results

The 50% status report is expected to show preliminary results toward addressing the problem you have selected, including preliminary classifier performance results.

Provide quantitative descriptions of how well the system components perform (*i.e.*, pre-processing, feature selection, etc.), and in the text of the report explain and interpret the results. It is insufficient to merely present a series of figures. Instead, talk the reader through the figures to ensure the reader is guided toward properly interpreting the figures and observing the key take-away points in the figures.

Some questions to think about are:

- Overall, how are results?
- For what cases do results look promising?
- For what cases so results fall below desired performance?

15% Conclusions / Challenges / Next Steps

Provide your overall assessment of the results for the pattern classification system you are building, including what you see as its strengths and weaknesses. Include descriptions of the challenges you are facing, as well as your plans for next steps.

Some questions to think about are:

- What about the system is working well, and is likely to remain largely unchanged?
- What about the system needs improvement, and will be modified?

5% Figures

Figures will be scored according to the expectations for the final report.

Support your textual content with figures. Support the description of your processing methods with flowcharts, and graphically depict performance (for example, do not present a list or table of AUCs for various conditions.) Your figures are expected to look professional. This means, at a minimum:

- Do not "Print Screen" a figure window, as this approach results in low quality images (doing so will result in point deductions) . Instead, export the figure to a graphics file and then import the high quality image into your document.
- Label all axes.
- Include a legend.
- Include a descriptive title.
- Ensure all text is large enough to be readable after the figure is imported into your document. 8-point font is generally accepted as the smallest usable font.
- To reiterate: Do not "Print Screen" a figure window. Instead, export the figure to a graphics file and then import the high quality image into your document.

10% References

References will be scored according to the expectations for the final report.

Include sufficient and appropriate references to archival material. This means books, journal papers, and conference papers. Do not cite wikipedia because it is not archival material - wikipedia changes, so someone who looks at the wikipedia page next week may not see the same information you saw yesterday. (Our class notes are also not archival material, and while I'm flattered to be cited it is not a reference that someone else reading your report will be able to access.) All classifiers and algorithms should be referenced. There are five textbooks I've made available to you this semester (Hastie *et al*, Izenman, and James *et al* as pdf files from the library, and Duda *et al*, and Bishop are on reserve at the library.) Everything we have discussed or will discuss in class is covered in at least one of those books.

Include reference(s)/citation(s) for:

1. Any statement or idea that is not common knowledge.
2. Any statement or idea that originated with someone else.
3. Any algorithmic approach you did not develop (*e.g.*, pre-processing, classifier, etc.).

Summarize or quote (and provide a citation to) relevant references rather than stating, "See page *x* in reference *y*."

References should be formatted according to standard formatting conventions, *i.e.*, use of **bold font**, *italic font*, and "quotes" to denote various pieces of information in the reference. The exact style for references may vary by discipline, but they all use formatting to varying degrees to clearly denote the different pieces of information in the reference, and the article title versus journal title or chapter title versus book title, in particular. One example of a commonly used formatting standard for references is the *IEEE Citation Index*, available at `http://www.ieee.org/documents/ieeecitationref.pdf`.