Pattern Classification and Recognition
ECE 681

Spring 2016
Course Project 25% Status Report Guidance

Due: 4:30PM Thursday, March 10[1]

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

Submit a **hard-copy** to the homework box in Teer (for us to mark-up and return to you).

Submit a **.pdf file** as an Attachment to the Assignment in Sakai (for us to refer to when evaluating future project report submissions).

Project 25% 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 25% status report is expected to document first steps toward addressing the problem you have selected, including preliminary data pre-processing and feature extraction/generation results, or preliminary theoretical developments.

15% Presentation Clarity

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.

There are no page requirements or limits – the 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).

Make use of the writing studio to help you improve your presentation.
`http://twp.duke.edu/twp-writing-studio`

It is acceptable to re-use text you originally wrote for your project proposal in your status report, though it is more likely you will revise and expand upon the text you have already written.

The expectations for the final report are higher than for the proposal, so start working now on revising and expanding your initial draft (your proposal) so the final report is complete and thorough. Someone reading your final report should have a thorough understanding of the problem you considered, the work you did, the results you found, and the conclusions or lessons learned that can be drawn from your work. The intermediate status reports are check-points along the way.

---

[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 mailbox in Hudson 130, or slid under my office door (Hudson 114).

Submitted by 4:30PM, Friday, March 11, 2016 = 1 day late. (10 points deducted from raw score)
Submitted by 4:30PM, Monday, March 21, 2016 = 2 days late. (20 points deducted from raw score)
Submitted by 4:30PM, Tuesday, March 22, 2016 = 3 days late. (30 points deducted from raw score)
Late submissions not accepted after 4:30PM, Tuesday, March 22, 2016.

Status reports submitted in person to me (Dr. Tantum), to my mailbox in Hudson 130, 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.

25% Problem Description

Provide background and context for the problem you have selected, and motivation for considering it. As you delve deeper into the problem, it is quite likely your own understanding of the problem background and context, in particular, will evolve. Your description of the problem should evolve as your understanding evolves and deepens.

The expectations for the Problem Description for the final report are higher than for the proposal, so start working now on revising and expanding your Problem Description so the description in your final report is 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.

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?

15% Data Description

Provide a description of the data you are considering, where it comes from, and why it is a good choice for the problem you have selected.

Present summary statistics (number of data points, proportion of data in each class, etc.) and visualizations of the data.

The expectations for the Data Description for the final report are higher than for the proposal, so start working now on revising and expanding your Data Description so the description in your final report is complete and thorough. Someone reading your report should have a good understanding of the data from your description of it. Visualizations are key; there is a reason behind the saying, "a picture is worth a thousand words."

Some questions to think about are:

- What data are you using?
- 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)?

15% Approach/Methods

Pre-Processing (10%)
Provide a description of and motivation for your preliminary steps toward pre-processing, including data normalization, and/or how you plan to generate and/or select features. The description should include necessary equations. For example, it is not enough to say ZmUv normalization was applied. Include the equation(2) to show how normalized features were obtained from the original (unnormalized) features.

The expectations for the Pre-Processing description for the final report are higher than for the proposal, so start working now on revising and expanding your Pre-Processing description so the description in your final report is 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. 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 (5%)
Provide a description of how you plan to approach developing classifier(s), including any design trade-off decisions you anticipate. Be clear about what classifier(s) you will code yourself, and for what classifier(s) you intend to use 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".

I recognize we have not yet delved in to all the specific classifiers we will cover this semester. Thus far, you have coded a K-Nearest Neighbors (KNN) classifier. In the remaining homeworks, you will code a Bayes Classifier, Fischer Linear Discriminant (FLD), Distance Likelihood Ratio Test (DLRT), Logistic Discriminant, Support Vector Machine (SVM), and Relevance Vector Machine (RVM), so you will have a variety of classifiers available in your personal toolbox to use and apply for your project.

The expectations for the Classifiers description for the final report are higher than for the proposal and this 25% status report. The final report is 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. Visualizations are key here, as well, and visualizations of decision surfaces are especially helpful.

Some questions to think about are:

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

5% Results

The 25% status report is expected to show results for first steps toward addressing the problem you have selected, including preliminary data pre-processing and/or feature extraction/generation results, or preliminary theoretical developments.

Provide quantitative descriptions of how well the system components performance (*i.e.*, pre-processing, feature selection, etc.), and in the text of the report interpret the results.

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?

10% Conclusions / Challenges / Next Steps

Provide your overall assessment of the the preliminary 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

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 should 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.

10% References

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.

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`.