Introduction to Machine Learning: Mini-Project 2

Project Requirements and Expectations

ECE 580
Spring 2022

Stacy Tantum, Ph.D.

Project Resources

You are not expected to write your own code to an SVM classifier or cross-validation. You may use toolboxes or packages that are available for your favorite computing platform (Matlab, Python, R, C++, etc.).

You are expected to implement an SVM classifier that provides a decision statistic (rather than just a binary decision).

Citations must be provided for every package and toolbox you leverage (multiple functions from within a single package or toolbox do not need their own citations).

If you choose to use packages or toolboxes, you are responsible for understanding how the package/toolbox implements the algorithm it provides, and being able to set options, parameters, etc. to achieve the outcomes we want. "But the function(s) I found didn't do the thing(s) you asked us to do" is not an acceptable reason for not achieving the project goals.

Collaboration

Even though you are each individually responsible for completing your own project and submitting your own Slidedoc describing your project, I strongly encourage you to collaborate extensively with others in the class, and especially others in your Peer Feedback cohort. You may interact with your classmates in much the same way you would interact with a team: share and debate ideas, collaborate on code and share your code, and compare and contrast results and interpretations of results, as a few examples.

Every student is responsible for completing their own project and submitting their own Slidedoc describing their project efforts and results. There are two motivations for requiring individual submissions:

1) it is to your benefit to understand every aspect of the project, and
2) it is to your benefit to be able to continue making progress toward completing the project even if another student's personal circumstances limit their ability to engage with the project for a time.

Recommended Project Milestones (Submission Deadline: **12:00NOON Wednesday 4/27**)

Week 1: (ends Friday 3/18)

- Ensure you can
  1) load/read the data
  2) plot (image) channel weights on the brain surface (you can create your own vector of 204 positive numbers to test)

Week 2: (ends Friday 3/25)

- Implement a two-class linear (no kernels!) SVM classifier
- Start Slidedoc

Week 3: (ends Friday 4/1)

- Ensure you can obtain decision statistics (in addition to a class decision) from your SVM classifier
- Implement two-level cross-validation
- Update and continue Slidedoc

ADVICE:

Do not assume you will be able to complete the project within only a week or two

Week 4: (ends Friday 4/8)

- Classify provided data (imagined movement and overt movement)
- Begin interpreting results (for discussion)
- Update and continue Slidedoc

Week 5: (ends Friday 4/15)

- Complete classifying data
- Interpret results and draw conclusions
- Update and finalize Slidedoc

Week 6: (ends Friday 4/22)

- Complete incomplete milestones

Before 12:00NOON Wednesday 4/27:

- Submit Slidedoc to **Gradescope**

Project Scoring

Project Criteria

Technical (35%)	Reporting (65%)	
	5%	Clarity and Organization
	5%	Visualizations
	10%	Introduction (Problem Description)
10%	10%	Mathematical Formulation
25%	10%	Experimental (Simulation) Results
	15%	Discussion / Conclusions
	3%	References
	3%	Collaboration Descriptions
	4%	Peer Feedback Session Participation

Scoring Criteria

	Technical ("doing the ML")	Reporting ("explaining/interpreting the ML")
9.5	Exceptional/insightful setup and implementation	Exceptional insight/explanation **and** full and complete narrative
9.0	Complete and correct setup and implementation	Full and complete narrative
8.5	Minor shortcomings in setup and implementation	Minor shortcomings in narrative
8.0	Shortcomings in setup and implementation	Shortcomings in narrative
7.0	Significant shortcomings in setup and implementation	Significant shortcomings in narrative
6.0	Major shortcomings in setup and implementation	Major shortcomings in narrative
5.0	Severe shortcomings in setup and implementation	Severe shortcomings in narrative
4.0	Little setup and implementation	Little narrative provided
3.0	No implementation	No narrative

Academic Integrity Expectations

It is expected that your project report represents your own ideas and your own understanding of the concepts written in your own words.

- The content is expected to be written in your own words
- You may consult resources and references provided that proper citations are provided for ideas you paraphrase
- You may quote resources and references in your report provided that proper quotations are used and proper citations are provided

It is not acceptable to:

- Copy content (text and/or images/figures) from class materials, including the project guidance, even if properly quoted and cited – **do not copy from class materials**
- Copy content (text and/or images/figures) from another resource/reference without both proper quotation and proper citation
- Paraphrase content (text and/or images/figures) from another resource/reference without proper citation
- Stitch together sequences of properly quoted and cited phrases and/or sentences to create passages that are not expressed in your own words

Project Criteria: Slidedoc format

A Slidedoc is a document that is more complete than a slide presentation, yet more concise than a conventional report.

My goal in going to this format for the project documentation is to streamline the "reporting back" process– there is only one document for you to prepare instead of two (a report and a separate presentation), and your effort for this single document is focused on efficiently communicating the salient points.

A Slidedoc that describes Slidedocs is available at: https://www.duarte.com/slidedocs/

There are no page requirements or limits for the Slidedoc; it should be parsimonious – as long as it needs to be to fully describe what you have done, but no longer than necessary.

Each component of the Slidedoc described here should be complete and thorough, so that someone reading your Slidedoc should have a good understanding of what you are describing solely from your descriptions of it.

References to relevant outside sources you used to support your project should be provided, and citations must be provided for any ideas, thoughts, statements, pictures, or figures that are not your own.

If a component is missing from the Slidedoc, then the corresponding score for that component is necessarily "No submission" (*i.e.*, 3.0 points).

Project Criteria: Slidedoc format (FAQs)

How are Slidedocs different from slide decks?

Slidedocs are intended to be read, whereas slide decks (presentations) are intended to be heard.

While a well-designed presentation (slide deck) is typically highly visual with very few words (often organized as bullet points), a well-designed Slidedoc includes prose (full sentences organized into short paragraphs). While the prose in a Slidedoc may be more concise (and scannable) than the prose in a long-form report, it is still prose, not bullet points. This means the reader should be able to fully understand the message the page conveys solely from reading the words on the page; the reader should not need to imagine additional dialogue (as would be provided by a presentation speaker) to fully understand the message the page conveys.

You can think about what the "speaker script" for a presentation slide might be and include that script as prose on the Slidedoc page. If you want to 'test' your page to see if it's a page from a Slidedoc or a page from a slide deck, read it out loud (only the words written on the page). Does it sound like a natural portion of a conversation? If it does, you have text for a good Slidedoc page. (Text alone does not necessarily make a good Slidedoc page; a good Slidedoc page typically also includes visual aid(s).) If, instead, the text on the page sounds like a series of disjoint statements, you do not (yet!) have text for a good Slidedoc page.

When I am reading the Slidedocs, I will read the words on the page; I will not imagine additional dialogue that may surround those written words if the Slidedoc were to be presented orally.

Project Criteria: Clarity and Organization

A slidedoc is a formal document. As such, writing (organization, sentence structure, spelling, etc.) matters, and the presentation clarity and organization score will reflect the quality of the written presentation. (I am not explicitly reading for grammar, spelling, etc., but I will notice if my comprehension is impeded by these elements and the Clarity and Organization score will reflect this.)

You are free to make use of writing support services on campus (which vary by the program in which you are enrolled) to help you improve your presentation.

Present a Slidedoc that is clearly written, easy to follow, and complete as a stand-alone document, as would a Slidedoc 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 mathematical concepts leveraged in this project, should be able to read your report and understand what you have done. Your presentation should describe the problem you are solving, describe your methods/approach to solving it, present your results, and present conclusions based on your results. This information should be presented in a logically organized, and sequential, way. Concepts should be defined or explained before they are used, and each page (slide) should have a key take-away point.

Project Criteria: Clarity and Organization (FAQs)

Why does clarity and organization matter when this isn't a writing class and I am not a writing teacher?

One of my jobs is to help you prepare for your professional career. Communication is a large component of many professional roles, so it is to your advantage to make the most of opportunities such as this to continue strengthening your communication skills. In many professional settings, strong communication skills are necessary for professional advancement, such as being appointed project lead or earning a promotion.

Specific pieces of advice:

- Define acronyms the first time they are used.
- Design slides titles to orient the reader to where they are in the Slidedoc.
- Make use of spell-checkers and grammar-checkers.

Project Criteria: Visualizations

Support your textual content with visualizations. It may be helpful to provide visualizations that illustrate the outcomes of various stages of the simulation process (i.e., original image, extracted block, sampled block, reconstructed block, filtered reconstructed block).

Your figures are expected to look professional. This means, at a minimum:

- Do not “Print Screen”, screen capture, or snip/clip a figure window, as this approach results in low quality images (doing so will result in point deductions) . Instead, export/save the figure as 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 readable font. (Making the figure window smaller prior to exporting the figure generally results in larger fonts in the exported figure.)
- Describe, in text, each figure on the same page (slide) it is presented in your Slidedoc.
- To reiterate: **Do not “Print Screen”, screen capture, or snip/clip a figure window.** Instead, export/ save the figure as a graphics file and then import the high-quality image into your document.

Project Criteria: Introduction

Provide an overview of the project in which you:

1) describe, in your own words, the nature of the project
2) describe, in your own words, the goals of the project (what is hoped to be achieved or gained at completion)
   (HINT: The goals go beyond inferring intended movement from EEG recordings.)
3) provide, in your own words, a summary of key results.

If you assimilate contextual information from reading other resources, then those resources are references for your problem description.

You can think of the introduction as your "elevator pitch" for the project. If you had 2 minutes to explain to someone what the project was about and what you learned from it (and you wanted them to be impressed enough to hand you a business card and ask you to call/email them tomorrow), what would you say?

Project Criteria: Mathematical Formulation

The audience for this project report is mathematically literate but is not familiar with support vector machines for classification. (This means you need to describe the underlying mathematics in your own words, even through our course notes and the project description/guidance given to the class provides this background information to you.)

- Describe, in your own words, how you formulated the classification problem.
- Describe, in your own words, how you applied an SVM to classify the EEG data.

It is insufficient to merely provide equations (the *what* or *how*). Also provide context that explains *why*, such as *why* is an SVM a good choice for this classification problem. The answers to *why* questions provide your motivation and justification for doing what you did, and your motivations and justifications are at least as important as *what* you did (the equations). You should be asking yourself "*why* am I doing this (and not something else, or doing nothing at all)?" for every step in the process

It is often helpful to include visualizations to support your textual descriptions.

Project Criteria: Experimental Results

Describe, in your own words, the simulation conditions and present the simulation 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 interpreting the figures in the way you intend for them to be interpreted and observing the key take-away points in the figures.

1) Specific experimental (simulation) results are required for each data set (overt and imagined movement):
   a) For the 1st-level cross-validation fold #1
      i. visualize the weights on the brain surface (using the provided show_chanWeight function)
      ii. show (plot) the weight for every channel
      iii. provide a list of the 5 dominant channels and their weights (this can be a separate table, or integrated into the visualization of all the weights in (ii) by clearly denoting the 5 dominant channels in the visualization and providing their weights)
   b) provide the ROC for each 1st-level cross-validation fold (6 ROCs), and the total cross-validated ROC on a single graph
      i. evaluate the cross-validation – are the individual fold ROCs representative of the total ROC?
   c) Provide accuracy for each 1st-level cross-validation fold assuming a threshold β=0 (6 accuracy values), and the total cross-validated accuracy
      i. evaluate the cross-validation – are the individual fold accuracies representative of the total accuracy?
2) Compare the regularization parameter values across the two data sets

You may also show any other interesting simulation results that you discover as part of your own explorations for this project. (This step is a necessary, but not sufficient, condition to earn a 9.5 on this section.)

Project Criteria: Discussion / Conclusions

Interpret, in your own words, your experimental results, including a comparison of the results obtained with the two data sets. For example, discuss *why* performance (accuracy) is similar or different, discuss *why* the number of support vectors is similar or different.

Topics may include, but are not limited to:

- Factors that may impact classification accuracy,
- Limits or problems with your approach,
- Possible improvements that can be made,
- Anything unique you have done to improve/validate your classifier's accuracy/efficiency.

Project Criteria: References

References / citations **must** be included

For example, support vector machines are well-established classifiers, so your description of the mathematical formulation must include citations to indicate to the reader that you are not the originator of this classifier. The references/citations must be books or journal articles. Websites are not suitable references, nor are our class lecture notes.

If you reproduce an image from another source (including websites) you must provide a citation for that image.

You must provide a citation for every toolbox or package you leverage. (We need to know what your code sources are.)

Citations must be provided for ideas that originate from other sources

Citations must be provided and quotation marks must be used for direct quotes from other sources

Citations must be provided for images that are not your own

It is not acceptable to take quotes or screenshots from other sources, including websites, without citation

It is not acceptable to take quotes or screenshots from our class course materials, including the project description, even with citation – do not copy from class materials

Project Criteria: Collaborations

Describe your collaborations with other people, including students in the class, while working on this project.

For example:

- Who did you share and debate ideas with while working on this project?
- Who did you share code with while working on this project?
- Who shared code with you while working on this project?
- Who did you compare results with while working on this project?
- Who did you help overcome an obstacle while working on this project?
- Who helped you overcome an obstacle while working on this project?

If you did not collaborate with anyone, state you did not collaborate.