Introduction to Machine Learning: Mini-Project 1

Project Requirements and Expectations

ECE 580
Spring 2022

Stacy Tantum, Ph.D.

Project Resources

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

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: **11:59PM Friday 3/4**)

Week 1: (ends Friday 1/28)

- Ensure you can
  1) load/read the images
  2) plot ("imshow") the images
  3) select a desired KxK block from an image
  4) sample pixels from a block (to simulate a compressed sensed or corrupted image)
- Start Slidedoc

Week 2: (ends Friday 2/4)

- Peer Feedback session #1 on Thursday 2/3
- Implement LASSO regression to estimate DCT coefficients for a single block
- Update and continue Slidedoc

Week 3: (ends Friday 2/11)

- Implement random subset cross-validation to choose the regularization parameter λ
  - Remember – λ may be different for each block!
- Begin simulations to explore the impact of the proportion of pixels that are sampled (S)
- Update and continue Slidedoc

Week 4: (ends Friday 2/18)

- Peer Feedback Session #2 on Tuesday 2/15
- Continue simulations to explore the impact of the proportions of pixels that are sampled (S)
- Begin interpreting results (for discussion)
- Update and continue Slidedoc

Week 5: (ends Friday 2/25)

- Complete experimental simulations
- Interpret results and draw conclusions
- Update and finalize Slidedoc

Before 11:59PM Friday 3/4:

- Submit Slidedoc to **Gradescope**

ADVICE:

The simulation results will take some time to run; do not assume you will be able to complete the project within only a week or two

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

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 the nature of the project
2) describe the goals of the project (what is hoped to be achieved or gained at completion) (HINT: The goals go beyond recovering missing pixels in an image. It also involves developing an understanding of the interplay between key parameters such as the fraction of pixels missing, the block size, the role of median filtering, etc.)
3) provide 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 compressed sensing image recovery. (This means you need to describe the compressed sensing process, including the underlying mathematics, in your own words, even through the project description/guidance given to the class provides this background information.)

- Describe how you formulated the image recovery problem as an under-determined linear system.
- Describe how you solve the regression problem with L_1-norm regularization to recover the image.
- Describe how you filtered the recovered image to improve image quality.

It is insufficient to merely provide equations (the *what* or *how*). Also provide context that explains *why*, such as *why* is the image recovery problem within the class of problems of solving underdetermined linear systems, *why* is L1-norm regression chosen as the regularization approach, and *why* median filtering is chosen over other filtering approaches such as frequency-selective (low-pass or high-pass) filtering. 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).

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

Project Criteria: Experimental Results

Describe 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.

Specific experimental (simulation) results are required:

1) For the small test image "fishing boat"
   a) Set the block size to 8x8 and consider five different sample sizes, i.e., S=10, 20, 30, 40, 50.
   b) Show a representative recovered image before and after median filtering for each sample size and provide the mean square error between the recovered image and the original image for each representative recovered image.
   c) Provide a plot comparing recovery error vs. number of samples with median filtering and without median filtering.
2) For the large test image "nature"
   a) Set the block size to 16x16 and consider five different sample sizes, i.e., S=10, 30, 50, 100, 150.
   b) Show a representative recovered image before and after median filtering for each sample size and provide the mean square error between the recovered image and the original image for each representative recovered image.
   c) Provide a plot comparing recovery error vs. number of samples with median filtering and without median filtering.

You may also show any other interesting intermediate and/or 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 your experimental results (provide explanations for *why* we see trends we see in the experimental results) and make recommendations regarding balancing trade-offs in the algorithm design (for example, how does the proportion of missing pixels influence the choice of block size, or should we always median filter the recovered image?).

Topics discussed in the conclusion may include, but are not limited to:

- Factors that may impact quality of the recovered image,
- Limits or problems with your approach,
- Possible improvements that can be made,
- Anything unique you have done to improve/validate your program's accuracy/efficiency.

Project Criteria: References

References / citations **must** be included

For example, LASSO regression and compressed sensing are well-established concepts, so your description of the mathematical formulation must include citations to indicate to the reader that you are not the originator of these concepts. 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 or our course materials) 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.)

References must be provided for outside sources you used to support your project

Citations must be provided for any images, ideas, or statements that are not your own

It is not acceptable to take quotes or screenshots from websites or our class course materials, including the project description, without citation

Project Criteria: Collaborations

Describe your collaborations with other 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 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.