

CS 288: Comments on Write-ups

In general, HW1 submissions were really good! However, I wrote up these comments to summarize the most common issues I saw. Because the homework process is designed to be as relevant as possible to the research paper process, most of these comments are also points that apply to submitting real research papers as well.

Criteria: What do I look for in a write-up? What do reviewers look for in a paper submission? Unfortunately, there's no rubric for peer review. However, a good paper should explore an interesting question, make clear claims, and support those claims with both empirical evidence and intuition. Good papers are well-written, rhetorically coherent, and neither too long nor too short. Readers like to (1) have a pleasant reading experience and (2) learn something.

(Re-)Writing: The single most common issue in submissions is that they appear to be written in one pass. There are many symptoms. Typos are one. Clumsy wording is another. Stream of consciousness organization is a third. My suggestion is that, unless you are very experienced in rewriting, you should read your paper aloud. You'll be amazed at how the errors and awkward phrasing jump out at you. Either way, make sure you rewrite and polish. At least run a spell-checker! Good writing is hugely important in any field; it's even more critical in NLP.

Rhetorical organization: Your paper should have coherent rhetorical structure. That often means having structural sections, like an abstract that tells the important points of the paper (what you tried, what the key results were). However, more importantly, it also means that you shouldn't ever just core dump your entire history of experiments on the reader. There should be high-level structure to the draft. There should also be clear low-level flow. Especially in a research paper, you should assume that your reader will start skimming unless you have made it clear why they are reading the current section. Are you explaining a key idea? Are you setting up the critical bit of error analysis? Read through your paper and ask yourself what purpose each paragraph serves. A related point: only show code when you have a point to prove with it that cannot be better proven in another way.

Target audience: This aspect is a little weird in a class context. You know that I know the KN formula, so what are you supposed to write? If you did considerable extensions (see below), you should concentrate on them. If you did the minimum, you should write a description which proves that you understood the material at a deep level, but which doesn't spend too much time on regurgitating unnecessary details. In general, you should think about who is reading the paper (who are the likely reviewers) and where they are coming from. Don't dwell too long on things they already know.

Related work: Especially in a real paper submission, you should cite your work. Here, that may or may not go beyond a cite to the textbook. However, listing references isn't the only task. You also need to situate your work with respect to those references. Are you describing something that is previously reported? Are you reimplementing it? Modifying it? How does your work relate to past publications? It should be totally clear what you claim your novel contributions are (in a basic submission here, there may not be any, but that should be clear, too).

Graphs and figures: In a talk, the situation is simple. Each figure or graph should make a single point, and the caption should state what that point is. Anything else is clutter. Tables with dozens of numbers are way too hard to process on the fly, and they should be distilled down to a

clear claim and proof. In a written paper, there's a place for having more data, so that your reader is free to draw their own conclusions. However, you should still beware of overloading your presentation. Full-page tables of specific numbers are rarely a good idea, and, when you do give a lot of information, make sure it's as well organized as possible. When in doubt, make sure each figure makes a simple claim and proves it. For papers, you also want your figures to be self-contained. Lots of reviewers just scan the figures and captions (!), so they should prove their important points on their own. Remember that your paper is a scientific argument, not a core dump. A last point about graphs and figures: make sure your text refers to them! I recommend making graphs and figures first, then writing to them, rather than writing a lot of text, then slapping in some figures. For example, sentences like "Figure 1 shows..." should usually come at the beginning of the relevant discussion, not the end.

Best results: You generally have some punchline result. This result is often your best accuracy on a standard test set, or something similar, but it may also be a trend you observed. Make sure your key results stand out. Bold them in a table, state them in your abstract or intro, and make it clear in the text what your top results are when you mention them. Lots of reviewers will evaluate your key results before reading the details of the paper. For example, it shouldn't be too hard for me to find your best WER in HW1.

Extension: Assignments for this class do not require original research. However, I always reserve the highest grades for submissions which show investigative curiosity, ambitious exploration beyond the minimum requirements. Did you detect interesting patterns in your results and chase them up? Did you choose to build something ambitious? Note that doing a good job of looking at your errors and thinking about the data is also a good component in this direction.

Formatting nitpicks: These are things I will never take off for, but which can annoy reviewers in a real paper submission setting. The underlying principle is "don't make your paper ugly." Some things to keep in mind in this direction:

- **Fonts:** Use the same font family throughout. If your text is in a serif font (Times, etc.), your headings and graphs should be, too. A common convention is serif for printed material, sans serif for slides, but not everyone follows this rule (and like any such convention, it's not always the right recommendation). Consistency is good in any case.
- **Color:** Using color to make your points is a great idea, and for slides it is all but required. However, whenever possible, your printed work should survive being rendered in black and white.
- **Significant digits:** If you tested on a few hundred sentences, do not report accuracies to machine precision. Usually three significant digits are enough, but there's no hard rule.
- **Conventions:** Quotes go "outside punctuation." Footnotes follow periods.¹ Neither of these conventions make sense from a nesting point of view; they are derived from typesetting behavior. Nominal Citation (2010) showed that author names should be inside the quotes only when parenthetical (Parenthetical Citation, 2010)