

CSC2541 for the Busy

I know that CSC2541 covers a lot of topics, there are a lot of readings, and many of the papers are pretty technical. If you have lots of deadlines coming up and don't have time to read everything, what should you focus on?

- The material that the course will build on is all covered in the NNTD notes. Reading the papers will let you understand the ideas at a deeper level and gain insight into the research context, but generally aren't needed in order to understand the following week's lecture. You can read them at your leisure.
- **Lecture 1:** Be familiar with the behavior of gradient descent on quadratics and be able to analyze it. Be able to reason about invariances and min-norm solutions. We'll use these ideas throughout the course, and you'll need them for the problem set.
- **Lecture 2:** Be comfortable working with JVPs/VJPs and formulating algorithms in terms of HVPs. Know what it means to linearize a network. This is the lecture you want to understand well in order to make sense of the rest of the course.
- **Lecture 3:** Be familiar with how the various curvature-related matrices (\mathbf{H} , \mathbf{G} , \mathbf{F}) relate to each other and when they'd be expected to match each other. Understand roughly what proximal optimization is doing. The details of Fisher information for exponential families are interesting, but not used all that much in the rest of the course. The PST will only be used in Lecture 4.
- **Lecture 4:** Be familiar with the various motivations for second-order optimization. The details of K-FAC won't be used much in the rest of the course.
- The remaining lectures will build on Lectures 1–4, but for the most part can be understood separately from each other. Lecture 8 builds on Lecture 7, Lecture 10 somewhat builds on Lecture 9, and as you'd expect, Lecture 12 directly builds on Lecture 11.