

# STAT 572 Course overview

Marina Meilă

Department of Statistics  
University of Washington

March 29, 2021

# Expectations

## Naming convention:

- “paper” = the research article you have chosen
- “report” = the paper you are going to write about this research article
- understand the paper
  - discuss with us if some proofs should be skipped
- place paper in context; do literature search, read background and follow-up papers if necessary
- implement the algorithm(s) or methods in the paper
  - you have to implement the paper’s methods yourself
  - OK to use existing code for methods you will compare with
  - OK (& encouraged) to use libraries for peripheral tasks (preprocessing, postprocessing, visualisation)
- reproduce the experiments in the paper
  - discuss with us which experiments, to what extent
- design new experiments (e.g to demonstrate additional properties, to compare with other methods)
- recommended: try a new programming language (python, C/C++) if you are serious about big data and machine learning

# Support and grading

- Instructors: Help, resources and feedback on writing papers, giving presentations, **answering questions**
- Mentors: Discuss paper-related questions
- Grading
  - based on effort – meeting the deadlines, incorporating feedback
  - grade is a form of feedback
  - Class vs. Prelim
    - 572 as training
    - Instructors are not examiners
    - Prelim outcomes

# Planning your project

- what the final writeup should contain
  - what proofs you are (not) responsible for
  - what you will implement
  - what extensions you will make to the assigned paper
  - what data sets you will use
  - what related literature you will discuss
  - the schedule
- 
- Remember this is the first draft. Will be revised as the term progresses.

## Plan – first draft

- the paper summary
- what parts of the paper you will implement
- what programming language will you use for implementation.  
if you think there is a reason to use code not written by you, which package and why (it is understood that you will use graphics/plotting software, and optimization software like LP, QP solvers, LBFGS as needed)
- are there any theoretical/methodological parts that you want to extend/develop?
- are there any problems with the paper that you will try to fix?
- data sets you will use (from the paper or from elsewhere) (optional, for now) a plan of the experimental evaluation
- extensions
- a schedule of the implementation stages by week (including week 2)
- a list of other related papers (background, followup research, competing methods, applications) that are relevant to your project

# The paper summary

Read the whole paper then ponder the answers to the following questions:

- ① What is the statistical problem that the paper is trying to solve? Challenges.
- ② What is the method/approach of the paper? Only general idea.
- ③ What are the scientific problems that can be solved by the new method?
- ④ What is the significance of the paper? How does it advance the state of the art?

Your summary should be organized around these questions (not necessarily in the given order).

Total pages: approximately 2–3