

CS 1671 / CS 2071 / ISSP 2071

Human Language Technologies

Session 9: Project match day, CRCD tutorial

Michael Miller Yoder

February 11, 2026

Quiz

- Go to **Quizzes > Quiz 02-11** on Canvas.
 - Covers Session 8: J+M 4-4.3
- You have until **1:10pm** to complete it
 - Must submit by 1:10pm
- Allowed resources
 - Textbook
 - Your notes (on a computer or physical)
 - Course slides and website
- Resources not allowed
 - Generative AI
 - Internet searches

Course logistics

- [Homework 1](#) is **due tomorrow, Thu Feb 12 at 11:59pm**
- After your project group is formed today:
 - Establish a communication channel (Discord, Teams through Pitt, email, Signal, WhatsApp, etc)
 - [Project proposal](#) due Feb 27, is the next deliverable

Overview: Project match day

- Project match process
- CRCRD resources available for the project
- Coding activity: custom features for logistic regression

Project match

- Go to the spot in the room with the project paper you are most interested in working on
 - We will likely do this for several rounds
- **Goal: groups of ~5 on projects**
- After groups are finalized, write your names on the back of the project paper

CRCO resources for the project

CRCO resources available for the project

- Storage space
 - 5 TB shared space for the whole class at `/ix1/cs1671-2026s`
- CLI for running scripts through the SLURM job scheduler
- Jupyter notebooks
 - Teach cluster
 - OnDemand

Logging into the CRCD with CLI

```
ssh <Pitt username>@h2p.crc.pitt.edu
```

- You will need to be on the [Pitt VPN](#) ([GlobalProtect app](#)) if you are not connected to WIRELESS-PITNET
- Your home directory only has 75 GB of storage!
- Check quota use with `crc-quota` command
- Feel free to store project data, code, etc at class storage space `/ix1/cs1671-2026s`
 - 5 TB available



Photo: Chuck Stout, from NARA & DVIDS

Running scripts with SLURM job scheduler

- You can run scripts (like Python scripts) on the CRCSD, just **don't run them directly on the nodes that you log into with ssh**
- Write a shell script with the commands you want and SLURM options at the top
- See the CRCSD documentation: <https://crc-pages.pitt.edu/user-manual/slurm/batch-jobs/>



Managing Python environments on the CRCSD

- See the CRCSD Python documentation: <https://crc-pages.pitt.edu/user-manual/applications/python/>
- First load a pre-installed Python version through Lmod (this loads environment variables)
 - Run `module spider python` to see options
 - Then `module load <module>`, e.g. `module load python/ondemand-jupyter-python3.11`
- Then create a conda environment (recommended over pip)
 - `conda create --prefix=/ix1/cs1671-2026s/<your_project>/<your_env>`
 - `source activate /ix1/cs1671-2026s/<your_project>/<your_env>`
 - `conda install <packages>`
- You can put `source activate /ix1/cs1671-2026s/<your_project>/<your_env>` in your shell script for SLURM

Jupyter options on CRCO

- There are two!
- JupyterHub on the teach cluster that we've been using in class is fine to use
 - If you need to install additional packages, please use your own Python environment, not the class environment
 - Feel free to select GPU options if needed
- If you need something for longer than 3 hours, see documentation on Open OnDemand (which also has an R portal): <https://crc-pages.pitt.edu/user-manual/web-portals/jupyter-ondemand/>
 - You request a server and they notify you when it's available
 - You can provide a path to a custom conda environment
 - Email Michael if you can't log in or have other issues



Coding activity: custom features for logistic regression

Notebook: custom features for logistic regression

1. Go to this [nbgitpuller link](#) (also available on course website)
2. Start a server with **TEACH – 6 CPUs, 48 GB**
3. Load custom environment at `/ix1/cs1671-2026s/class_env`
4. Open `session8_logistic_regression_features.ipynb`