Our code structure is based on the instanovo package found at
https://github.com/instadeepai/InstaNovo/tree/main

Our experiments were run by modifying their github code to create instanovo+ marg, which is based on instanovo+. The instanovo folder was not changed (besides configuration parameters), and is based on the original paper. The instanovo_marg folder contains our modifications, and is used to train and run instanovo+ marg.

Our results can be found in the results folder. Note .out files must be opened using a text editor

Setup:
1. install uv,
For linux/mac:
curl -LsSf https://astral.sh/uv/install.sh | sh

For windows:
powershell -c "irm https://astral.sh/uv/install.ps1 | iex"

2. Clone our repository (or download the zip file submitted on learn), then enter the folder
git clone https://github.com/seantanger/InstaNovo.git
cd InstaNovo

3. Create a virtual environment if there isn't one in the directory, .venv, and activate it
uv venv .venv # Create
.\.venv\Scripts\activate # For windows
source .venv/bin/activate # For mac

4. Install dependencies
CPU Only:
uv sync --extra cpu
uv run pre-commit install

With GPU:
uv sync --extra cu124
uv run pre-commit install

5. Check if instanovo is installed
instanovo version

Training:
- Download proteomic tools dataset by running the get_data_proteome_tools.py script
- Generate the prior distribution based on the training dataset for instanovo+ marg by running the marginal_distribution.py script found in instanovo_marg/utils/marginal_distribution.py

- Set training parameters in instanovo/configs/instanovoplus (for Instanovo+), or in instanogo_marg/configs/instanovoplus_marg (for Instanovo marg +)

- Train Instanovo+ using terminal command (or bash script train_diffusion.sh)
  Instanovo diffusion train

- Train Instanovo marg+ by running the train_diffusion_marg.py script

Models are saved in a checkpoint folder every 25000 training updates.

Prediction:

- To run predictions using instanovo+, run the bash script predict_diffusion.sh, or terminal command “instanovo diffusion predict”
  Set --instanovo-plus-model to the path of ckpt model
  Set --data-path to the path of the test data to predict on
  Set --output-path to the path of your output data
  Add --no-refinement --evaluation

- To run predictions using instanovo marg+, run the python script predict_diffusion_marg.py.

Set the same parameters for instanovo+ at the bottom of the script, where the function run_diffusion_predict() is called.

- To run predictions using instanovo transformer, run the bash script predict_transformer.sho or terminal command “instanovo transformer predict”

Set --instanovo-model to the path of pretrained instanovo model (can be downloaded from github)
Set --data-path to the path of the test data to predict on
Set --output-path to the path of your output data
Add --evaluation