CSE 151B Deep Learning
UCSD, Spring 2021

Project Final Report

Due 11:59 p.m, June 6th, 2021

Please submit your report as a single PDF file using the NeurIPS 2021 template:

- Comment out `\usepackage{neurips_2021}`(Line 8) and Abstract (Line 75)
- Uncomment `\usepackage[preprint]{neurips_2021}` (Line 12)
- Change title to **CSE 151B Project Final Report** (Line 30)
- Change author name and email to your team members' names and emails (Line 40)
- Follow the format instructions to write your report, especially Figures and Tables

You report will contain the following components with a minimum 4 pages (not including references). Some of the questions are the same as the milestone. You can reuse the materials from your milestone report. You are also welcome to update your answers given the new things you have learned since the milestone. Please copy the section titles, the problem numbers and write the details under each section title. Do **not** copy the questions themselves. Include a **github** repository link for your solution. You are allowed to make the repository public at this point.

Your report will be judged by (1) how well it is following the format requirements (2) how much detail it contains about your project (3) how creative/innovative/thoughtful your solution is for the project.

1 Task Description and Background

Problem A [0.5 points]: Describe in your own words what the deep learning task is and why it is important. Provide some real-world examples where solving this task can have great impact on our daily life and the society.

Problem B [0.5 points]: Use Google Scholar or other internet resources to research on this task. What type of methods have been examined before? Include some references and discuss their ideas in a few sentences. You can use Bibtex to manage your bibliography.

Problem C [1 points]: Define the input and output in mathematical language and formulate your prediction task. From the abstraction, do you think your model can potentially solve other tasks beyond this project? If so, list some examples and explain your rationale.

2 Exploratory Data Analysis

Perform exploratory data analysis and report your findings with texts/tables/figures. If you include more exploratory analysis beyond the listed questions that provides insights into the data, you will receive bonus points.

Problem A [0.5 points]: Run the provided Jupyter notebook for loading the data. Describe the details of this dataset. Your description should answer the following questions:

- what is the train/test data size?
- how many dimensions of inputs/outputs in the raw data?
- what are the meanings of these input/output dimensions?
- what does one data sample looks like?

Problem B [0.5 points]: Perform statistical analysis to understand the properties of the data. Your analysis should at least answer the following questions.

- what is the distribution of input positions/velocity (magnitude) for all agents
- what is the distribution of output positions/velocity (magnitude) for all agents
- what is the distribution of positions/velocity (magnitude) for the target agent

Problem C [1 points]: Process the data to prepare for the prediction task. Describe the steps that you have taken to process the data. Your description should at least answer the following questions.

- Did you use any feature engineering? If yes, how did you design your features? Explain your rationale.
- How did you normalize your data? Why did you choose this normalization scheme?
- Did you use the lane information provided in the dataset. If yes, how did you exploit this information.

3 Deep Learning Model

Problem A [1 points]: Describe the deep learning pipeline for your prediction task and answer the following questions.

- What are the input/output that you end up using for prediction after pre-processing?

- What is your loss function? If you have multiple alternatives, discuss your ideas and observations.
- How did you decide on the deep learning model? Explain the rationale given the input/output.

Problem B [1 points]: Describe all the models you have tried to make predictions. You should always start with simple models (such as Linear Regression) and gradually increase the complexity of your model.

- Use an itemized list to briefly summarize each of the models, their architecture, and parameters, and provide the correct reference if possible.
- If you end up designing your own model architecture, include a picture/sketch of your model architecture. Explain why you choose such a model.
- Describe different regularization techniques that you have used such as dropout and max-pooling in your model.

You can also use mathematical equations to explain your prediction logic.

4 Experiment Design

Problem A [1 points]: Describe how you set up the training and testing design for deep learning. Answer the following questions:

- What computational platform/GPU did you use for training/testing?
- How did you split your training and validation set?
- What is your optimizer? How did you tune your learning rate, learning rate decay, momentum and other parameters?
- How did you make multistep (30 step) prediction for each target agent?
- How many epoch did you use? What is your batch-size? How long does it take to train your model for one epoch (going through the entire training data set once)?

Explain why you made these design choices. Was it motivated by your past experience? Or was it due to the limitation from your computational platform? You are welcome to use screenshots or provide code snippets to explain your design.

5 Experiment Results

Problem A [1 points]: Select a few representative models of yours and report the following results by comparing different models.

- Use a table to compare the performances of different model designs. What conclusions can you draw from this table.
- Provide an estimate of training time (in flops or minutes) for different models. What did you do to improve the training speed?
- Count and report the number of parameters in different models.

Problem B [1 points]: Play with different designs of your model and experiments and report the following for your best-performing design:

- Visualize the training/validation loss (RMSE) value over training steps (You should expect to see an exponential decay).
- Randomly sample a few training samples after the training has finished. Visualize the ground truth and your predictions.
- Your current ranking on the leaderboard and your final test RMSE.

6 Discussion and Future Work

Problem A [1 points]: Analyze the results and identify the lessons/issues that you have learned so far. Briefly answering the following questions.

- What do you think is the most effective feature engineering strategy?
- What techniques (data visualization/model design/hyper-parameter tuning) did you find most helpful in improving your score?
- What was your biggest bottlenect in this project?
- How would you advise a deep learning beginner in terms of designing deep learning models for similar prediction tasks.
- If you had more resources, what other ideas would you like to explore?