Week 11 Assignment

1.

Week_1 quastion：如果 data set 裡的 data,出現少數幾點與其他 data 相比偏離很大，使用 MSE loss 可能會導致因這幾點的影響過大，而降低對其他多數點的預測。這種情況怎麼調整 loss function 或 data set？

Reference：

Huber loss — https://en.wikipedia.org/wiki/Huber_loss.

Tukey's biweight — https://www.r-bloggers.com/2021/04/what-is-the-tukey-loss-function/.

IRLS（Iteratively Reweighted Least Squares）— https://en.wikipedia.org/wiki/Iteratively_reweighted_least_squares

Week_2 question：What is the good way to determine the learning rate of the optimizer？

Reference：

Leslie N. Smith — *Cyclical Learning Rates for Training Neural Networks*（LR range test, CLR）。https://arxiv.org/abs/1506.01186

Week_3 queation：For a neural network to approximate a function f and its derivative f' at the same time, I find it's hard to train. I first choose the loss to be MSE loss of f + MSE loss of f' and the result didn't go well. Then I think that if the MSE loss of f' is weighted then it will be better, and it did be better, but it's not good enough. I don't know if I should choose another weight or there is other way to make it better?

Reference：

Sobolev Training for Neural Networks（Czarnecki et al., NeurIPS 2017）— 關於用 target derivatives 的訓練方法與效果。https://arxiv.org/abs/1706.04859

GradNorm: Gradient Normalization for Adaptive Loss Balancing — 自動平衡多任務梯度。https://proceedings.mlr.press/v80/chen18a.html

Physics-informed neural networks（Raissi et al.）— PINNs 的訓練策略與實務建議。https://arxiv.org/abs/1711.10561

Week_4 question：I found that the choice of the training set, test set and validation set will affect the result. Since the form of the data in the xml file, I can't just use the first 80 percent of the data to do the choice to get a good result, so I choose randomly. I'm wondering that if there is any other way to choose these data set？

Reference：scikit-learn 的 train_test_split、StratifiedShuffleSplit、GroupKFold 等文件。https://scikit-learn.org/

Week_5 question：In Gaussian discriminant analysis, Σ_k needs to be invertible, but if $n_k<n$, then Σ_k can be not invertible. How to deal with this situation? Is regularization the solution?

Reference：Friedman, *Regularized Discriminant Analysis* (1989)
https://sci2s.ugr.es/keel/pdf/algorithm/articulo/friedman1989.pdf

Ledoit & Wolf, *A well-conditioned estimator for large-dimensional covariance matrices* (2004)

https://perso.ens-lyon.fr/patrick.flandrin/LedoitWolf_JMA2004.pdf

Week_6 question：I found that the speed of running the regression code in this assignment is way slower than the classification code. I suspect that it is because of the neural network I use in the regression code since the code which I also use neural network in previous assignments is also slow. How to make it faster?

Reference：https://developer.nvidia.com/blog/profiling-and-optimizing-deep-neural-networks-with-dlprof-and-pyprof

Week_7 question：In the first of the score-based diffusion generative model, if I change the noise distribution from Gaussian to a different probability distribution noise, then how would this change the reverse process?

Reference：關於以 Lévy/non-Gaussian 驅動的 score-based 模型（time-reversal 與 jump）：E. Yoon et al., *Score-Based Generative Models with Lévy*

Processes（2022/2023）

https://papers.nips.cc/paper_files/paper/2023/file/8011b23e1dc3f57e1b6211ccad498919-Paper-Conference.pdf

Week_8 question：If the drift and the diffusion term in SDE don't satisfy certain regularity condition, than what will happen, why we need this condition？

Reference：https://www.uni-ulm.de/fileadmin/website_uni_ulm/mawi.inst.020/kunze/SDE/sde_skript.pdf

Week_10 question：關於 SDE，Ito's integral 和 Stratonovich's integral 有什麼區別？我查到的資料說後者常用在物理領域，選擇兩者間的哪一個來使用的考量是什麼？

Reference：介紹性筆記：*Itô and Stratonovich; a guide for the perplexed*
https://oatml.cs.ox.ac.uk/blog/2022/03/22/ito-strat.html?utm_source=chatgpt.com

綜述：*Itô versus Stratonovich: 30 years later*

https://www.researchgate.net/publication/263988676_ITO_VERSUS_STRATONOVICH_30_YEARS_LATER