UDACITY

PROJECT

Machine Learning Capstone Project

A part of the Machine Learning Engineer Nanodegree Program

PROJECT REVIEW

CODE REVIEW

NOTES

SHARE YOUR ACCOMPLISHMENT!

Meets Specifications

Dear student

Great job updating your report! Allow me to be the first to congratulate you on completing the Machine Learning Nano Degree. I've carefully weighed the previous review so that we're not asking for new things from one review to the next. However, I feel that you've met or exceeded the specifications for this course and your deeper dive into supervised learning has been quite successful. Again, congratulations on passing and I wish you all the best of luck with your future programming endeavors.

Cheers!

Definition

Student provides a high-level overview of the project in layman's terms. Background information such as the problem domain, the project origin, and related data sets or input data is given.

Nice overview of the problem domain! I love the focus on the real-world impact of the application.

Suggested:

- It's a good idea to cite some of the studies where the machine learning techniques that you're using were pioneered. This shows that you really know the field and it gives credit to the inventors.

The problem which needs to be solved is clearly defined. A strategy for solving the problem, including discussion of the expected solution, has been made.

You've done a great job restating the problem clearly!

Suggested:

- This is a good point to begin to justify why your solution is a good 'fit' for the problem. If you were submitting this to a journal for peer review, you'd want to keep the readers focused on what you want them to think about. If they get distracted, they can ask for random things in subsequent revisions (which can significantly drag out the process and lead to arguments).

Metrics used to measure performance of a model or result are clearly defined. Metrics are justified based on the characteristics of the problem.

Analysis

If a dataset is present, features and calculated statistics relevant to the problem have been reported and discussed, along with a sampling of the data. In lieu of a dataset, a thorough description of the input space or input data has been made. Abnormalities or characteristics about the data or input that need to be addressed have been

identified.

Very nice job! You've made a thorough description of the dataset here.

A visualization has been provided that summarizes or extracts a relevant characteristic or feature about the dataset or input data with thorough discussion. Visual cues are clearly defined.

Algorithms and techniques used in the project are thoroughly discussed and properly justified based on the characteristics of the problem.

Nice description of EM!

Student clearly defines a benchmark result or threshold for comparing performances of solutions obtained.

This is very clear now. Well done!

Methodology

All preprocessing steps have been clearly documented. Abnormalities or characteristics about the data or input that needed to be addressed have been corrected. If no data preprocessing is necessary, it has been clearly justified.

The process for which metrics, algorithms, and techniques were implemented with the given datasets or input data has been thoroughly documented. Complications that occurred during the coding process are discussed.

Great job! I think that a skilled programmer should be able to (more or less) reproduce your results using only the report at this point.

The process of improving upon the algorithms and techniques used is clearly documented. Both the initial and final solutions are reported, along with intermediate solutions, if necessary.

The steps you took towards refinement are documented as are the results.

Results

The final model's qualities — such as parameters — are evaluated in detail. Some type of analysis is used to validate the robustness of the model's solution.

> Comparing it with the benchmarks, the Gaussian Mixture performed better than both of the them, however not significantly better than GaussianNB which scored an AuC 0.68.
> I'd like to point out an advantage over the GaussianNB that we don't need the labels since it's an Unsupervised Learning technique (under certain domains is hard to acquire), even though we used the labels to build the model.

Nice analysis! This could certainly make for a more flexible framework.

The final results are compared to the benchmark result or threshold with some type of statistical analysis. Justification is made as to whether the final model and solution is significant enough to have adequately solved the problem.

Conclusion

A visualization has been provided that emphasizes an important quality about the project with thorough discussion. Visual cues are clearly defined.

> I had to study a lot about probability distribution, linear algebra and some calculus stuff. The first implementation was taking minutes to train the model. The second implementation was based on the scikit-learn Gaussian Mixture. This one was taking seconds to train the model. So, it was a great improvement.

Speed is definitely an advantage of the SKLearn library if you're doing machine learning in Python.

Student adequately summarizes the end-to-end problem solution and discusses one or two particular aspects of the project they found interesting or difficult.

Nice overview of your solution! Keep in mind that you can brag a bit about the novel/interesting parts in your solution.

Discussion is made as to how one aspect of the implementation could be improved. Potential solutions resulting from these improvements are considered and compared/contrasted to the current solution.

Quality

Project report follows a well-organized structure and would be readily understood by its intended audience. Each section is written in a clear, concise and specific manner. Few grammatical and spelling mistakes are present. All resources used to complete the project are cited and referenced.

The template is followed reasonably closely, and you've included all the required components. Additionally, justifications for each decision are made and this demonstrates a sufficient level of critical thinking for academic publication.

Code is formatted neatly with comments that effectively explain complex implementations. Output produces similar results and solutions as to those discussed in the project.

DOWNLOAD PROJECT

RETURN TO PATH

Student FAQ