IE 582 Statistical Learning for Data Mining
Homework 1, Due November 15, 2024

1. Introduction:

With the evolution of high-frequency communication systems, such as those required for 5G technology, the design of effective antennas has become crucial. Antennas in these systems must meet strict performance standards to ensure efficient transmission and reception of signals across specified frequency bands. However, accurately predicting antenna characteristics like the S-parameters, particularly S_{11}, which measures reflection and is crucial for assessing how well the antenna radiates at desired frequencies, is a computationally intensive task.

Traditional electromagnetic (EM) simulations, while precise, require significant computational resources and time, especially when exploring a wide range of geometric configurations and material properties. So it is very hard to perform trial and error to find out the antenna design that meets the user requirements. A sample antenna design problem from [1] is illustrated in Figure 1 and associated design parameters (i.e. geometry) parameters is given in Table 1.

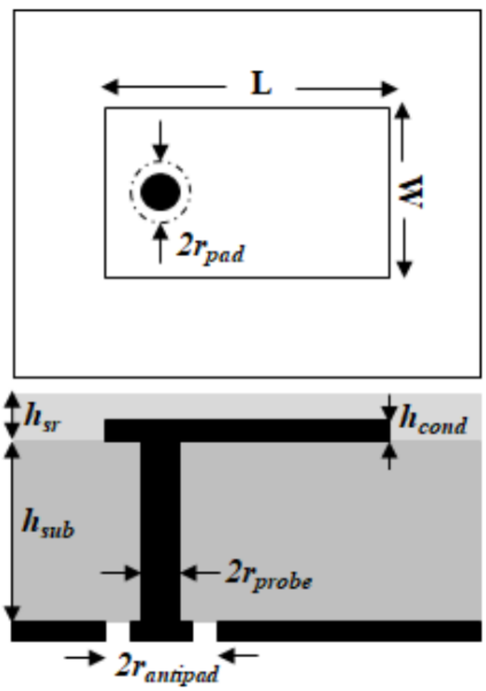

Figure 1. Top and cross section views of the microstrip patch antenna [1].

The designer is mainly interested in finding out S_{11} (return loss) values over multiple frequencies. In order to understand S_{11} , a potential solution is to sample some geometry points using the ranges in Table 1 and use an electromagnetic solver (i.e. simulation) to obtain the respective behavior. The designs considered in this homework are simulated within the frequency band of 23-33 GHz by [1] using Ansys HFSS (High Frequency Structural Simulator) [2] and complex S_{11} values (in terms of real and imaginary parts), which were collected at 201 frequency points to comprehensively evaluate antenna performance, become the output dataset.

Table 1. Design parameters and their ranges [1]

Parameter	Range
Patch Length (L)	1.8–5.2 mm
Patch Width (W)	1.8–5.2 mm
Probe radius	0.015–0.05 mm
Metal thickness (h_{cond})	0.01–0.04 mm
Substrate thickness (h_{sub})	0.1–0.8 mm
Solder resist thickness (h_{sr})	0.02–0.1 mm
Pad-probe radius difference	0–0.025 mm
Anti-pad-pad radius difference	0.025–0.1 mm
Probe position/patch length ratio	0.05–0.45
Substrate dielectric constant	2–5
Solder resist dielectric constant	2–5

Since the simulations are time consuming (computationally expensive), finding ideal design may not be easy. Due to such challenges, machine learning methods are increasingly used to understand the relation between input space (i.e. geometry) and output (i.e. real and imaginary parts of S_{11} values). In this context, dimensionality reduction techniques, including Principal Component Analysis (PCA) or supervised learning techniques such as linear regression offer powerful tools for data-driven analysis and prediction in antenna design applications. This assignment leverages simulated antenna design data to explore how these methods can capture and simplify complex relationships within the S-parameter dataset. The objective is to enhance understanding the relation between the design parameter and S_{11} space.

2. Data:

You are provided the data for 385 simulation runs which is generated randomly using the range of the design parameters given in Table 1. The file name is “hw1_input.csv” in comma separated values (*.csv) format where each row represents the geometry and each column is the design parameter (11 parameters in total).

There are two files for the output, namely “hw1_real.csv” and “hw1_img.csv” , providing the information for the real and imaginary part of the S-parameter respectively. The rows represent the design parameter with the same order of “hw1_input.csv” and the columns are representing the frequency levels where S-parameter is evaluated.

All files are provided as a *.zip file on Moodle. Here is [the direct link.]()

3. Research Questions:

This assignment focuses on exploring the relationships between the design parameters of an antenna and its electromagnetic performance using machine learning techniques. The goal is to employ dimensionality reduction and predictive modeling approaches to address the following research questions:

Note that electromagnetic performance is sometime quantified by the magnitude of the S-parameters. You may want to perform a brief research to understand what an S-parameter is. The magnitude profiles for the first 6 designs in the given dataset are illustrated in Figure 2.

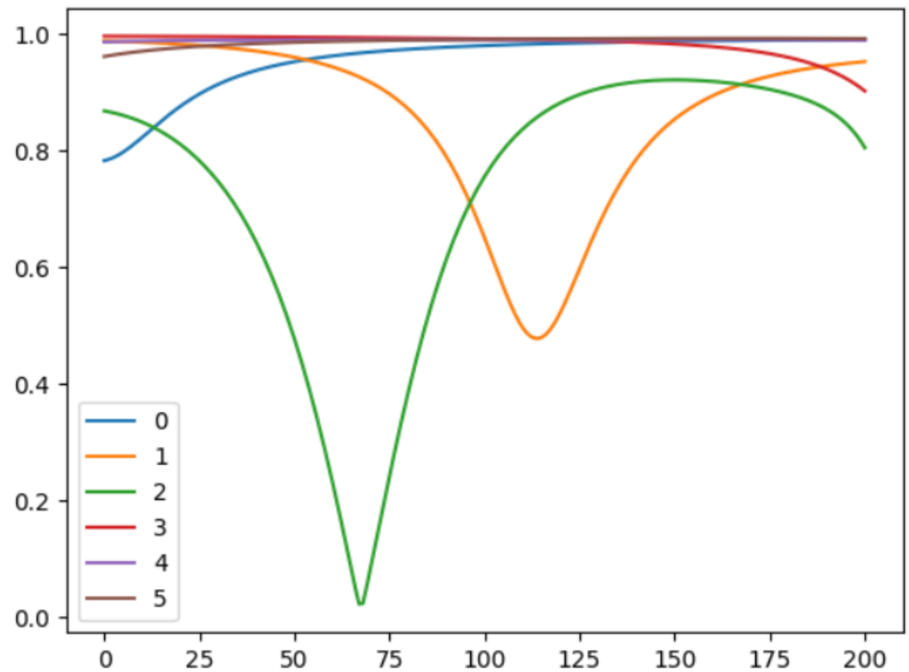

Figure 2. S_{11} magnitudes over 201 frequency indices for the first 6 designs in the given dataset. For example, the green design (design indexed at 2, which is the third row) shows the S_{11} value approaching zero around the frequency at index 70. Smaller magnitudes indicate less signal loss, meaning a stronger reflection performance. In the case of S_{11}, a smaller magnitude implies that the signal can be received more effectively around that frequency.

3.1. Dimensionality Reduction with PCA:

- Can we reduce the complexity of the design space by using Principal Component Analysis (PCA) to identify key parameters that most influence the S_{11} response of the antenna?
- How much of the total variance in the design parameter space can be explained by the principal components?
- What insights can we draw from the PCA regarding the relationship between geometry and electromagnetic behavior?

3.2. Regression Modeling for S_{11}

- Given that S_{11} parameters are evaluated at 201 frequency points, predicting them simultaneously through multitarget regression may be computationally intensive and beyond the scope of the content covered so far. Instead, can we simplify the regression task by focusing on predicting S_{11} at a few key frequency points? For instance, selecting frequencies of interest where the behavior is most critical (e.g., resonance frequencies) can reduce the complexity of the task.
- How effective are linear regression models in predicting the real and imaginary components of S_{11} at these selected frequency points, based on the geometric parameters of the antenna design?
- What patterns emerge when linear regression is applied to individual frequency points, and do these patterns suggest any broader trends in the design space?

3.3. Model Performance and Interpretability:

- How do PCA and regression models compare in terms of their ability to simplify and predict the antenna's performance?
- What are the potential limitations of these models, and how could they be improved to more accurately represent complex, nonlinear electromagnetic behavior?

4. Report & Code Documentation:

Combine your results and visual aids into a comprehensive report. Your report should:

- Discuss your methodologies and findings.
- Draw conclusions based on the patterns observed and their possible real-world implications.
- Offer insights into any challenges faced and how you overcame them.

You can work with any language you want (i.e. R, Python, Julia, Matlab and etc.). You are expected to use GitHub Classroom and present your work as an html file (i.e. web page) on your progress journals. There are alternative ways to generate an html page for you work:

- A Jupyter Notebook including your codes and comments. This works for R and Python, to enable using R scripts in notebooks, please check:
    - https://docs.anaconda.com/anaconda/navigator/tutorials/r-lang/
    - https://medium.com/@kyleake/how-to-install-r-in-jupyter-with-irkernel-in-3-steps917519326e41

  Things are little easier if you install Anaconda (https://www.anaconda.com/). Please export your work to an html file. Please provide your *. ipynb file in your repository and a link to this file in your html report will help us a lot.

- A Markdown html document. This can be created using RMarkdown for R and Python. Markdown for Python

Note that html pages are just to describe how you approach to the exercises in the homework. They should include your codes. You are also required to provide your R/Python codes separately in the repository so that anybody can run it with minimal change in the code. This can be presented as the script file itself or your notebook file (the one with *.ipynb file extension).

Academic Integrity

The last and the most important thing to mention is that academic integrity is expected! Do not share your code (except the one in your progress journals). You are always free to discuss about tasks but your work must be implemented by yourself.

Homework assignments in this course permits the use of GenAI tools. Any such use must be appropriately acknowledged and cited. It is each student’s responsibility to assess the validity and applicability of any GenAI output that is submitted; you bear the final responsibility.

For all other forms of assignments, quizzes and exams use of GenAI tools is disallowed.

5. Conclusion

Data mining is both an art and science. As you journey through this homework, remember that the process is as valuable as the outcome. Your ability to connect the dots between seemingly unrelated pieces of information will shape your growth in the realm of data science.

We're eager to see the unique insights each of you will bring to this project, especially given the diverse academic backgrounds of our cohort. Best of luck, and happy mining!

References

[1] E. S. Saçın and A. C. Durgun, "Neural Network Modeling of Antennas on Package for 5G Applications," 2023 17th European Conference on Antennas and Propagation (EuCAP), Florence, Italy, 2023, pp. 1-5, doi: 10.23919/EuCAP57121.2023.10133407.

[2] ANSYS HFSS software. http://www.ansoft.com/products/hf/hfss/