1 Introduction

Modelling turbulent transport in fusion plasmas from first principals at this time is at best impractical, and at worst, impossible.

As such, data-driven methods used to construct scaling laws have taken preference. The most notable of which is the IPB98 scaling law:

$$\tau_{\mathrm{E,th}}^{\mathrm{ELMy}} = 0.0562 I^{0.93} B^{0.15} P^{-0.69} n^{0.41} M^{0.19} R^{1.97} \varepsilon^{0.58} \kappa_a^{0.78} \tag{1}$$

Since then, over two decades of additional data contributed from a plethora of machines has shown a clear decreasing trend in the dependence of the major radius, with the most recent iteration showing:

$$\begin{aligned}\tau_{\mathrm{E,th}} = 0.053 I_{\mathrm{p}}^{0.98\pm0.19} B_{\mathrm{t}}^{0.22\pm0.18} \\ \times \bar{n}_{\mathrm{e}}^{0.24\pm0.11} P_{\mathrm{1,th}}^{-0.669\pm0.059} R_{\mathrm{geo}}^{1.71\pm0.32} \\ \times (1+\delta)^{0.36\pm0.39} \kappa_{\mathrm{a}}^{0.80\pm0.38} \epsilon^{0.35\pm0.66} M_{\mathrm{eff}}^{0.20\pm0.17}\end{aligned} \tag{2}$$

Further exaggerated when we consider an ITER-like subset of the data:

$$\begin{aligned}\tau_{\mathrm{E,th}} = 0.067 I_{\mathrm{p}}^{1.29\pm0.16} B_{\mathrm{t}}^{-0.13\pm0.17} \\ \times \bar{n}_{\mathrm{e}}^{0.147\pm0.097} P_{\mathrm{1,th}}^{-0.644\pm0.061} R_{\mathrm{geo}}^{1.19\pm0.27} \\ \times (1+\delta)^{0.56\pm0.36} \kappa_{\mathrm{a}}^{0.67\pm0.63} M_{\mathrm{eff}}^{0.30\pm0.16}\end{aligned} \tag{3}$$

The design of next-generation machines will be guided by these scaling laws, therefore warranting further investigation.

2 Problem

Given the following linear regression model:

$$\tau_{\mathrm{E,th}}^{\mathrm{ELMy}} = \alpha_0 I^{\alpha_I} B^{\alpha_B} P^{\alpha_P} n^{\alpha_n} M^{\alpha_M} R^{\alpha_R} \varepsilon^{\alpha_\epsilon} \kappa_a^{\alpha_\kappa} \tag{4}$$

Consisting of nine correlated variables, such that, when the model is applied to two datasets, A and B, one of which is an updated version of the first:

- $A \cap B = A$
- $B / A = U$

We obtain a significant reduction in a coefficient of interest. The task is to determine the smallest possible set of data points in U which contributes the most to this reduction.

- This is a nine dimensional optimisation problem. By reading into the literature surrounding multidimensional optimisation problems, you need to decide on a loss function that suits the problem at hand, and then using a given optimisation method, attempt to determine the global minimum.
- Each method of optimisation has associated with it, either a simplicity in concept with a higher computational complexity, or vice versa. You can each choose one of the following optimisation methods:

- Random search followed by a gradient descent in a selected area of the optimisation manifold
- Monte-Carlo Markov chain
- Simplex model

▸ What do the nine values we obtain tell us about the correlation/relation between the variables, if any?

▸ Do the nine values we obtain indicate whether or not optimising over all nine dimensions are strictly necessary?

3 Additional work

Once we have determined the smallest set of datapoints $D \in U$, responsible for the biggest influence in the reduction of the major radius dependence, we would further like to understand why.

Dimensional reduction is a well documented field in machine learning, with a multitude of pathways to follow, you will read around the subject and determine which algorithm would be best suited to our problem, and then apply the algorithm (potentially multiple) to determine what features of D are influencing the reduction.

Potential dimensionality reduction techniques:

▸ subspace clustering
  - bottom up approach
  - top down approach

▸ Neural nets

▸ Autoencoders

▸ UMAP

▸ T-SNE