

The second competition (Regression problem) has been posted on Kaggle. You can access it via the following link:

<https://www.kaggle.com/t/3b21ee7c45b049d0a273100fbd355aeb>

You are expected to participate in the competition individually and use your proper full name. This is important because many people have similar first names. I would remove any entry with an indistinguishable name/username (like warrior, beast, blackbeauty, etc.).

Please submit a PowerPoint describing the data preparation and algorithms that you attempted and specially identify the things that work for you. You may eventually submit 100s of submissions but make sure that your presentation contains your best results for the following algorithms:

- Regression Tree
- Random Forest Regressor
- Gradient Boosting Regressor
- Linear Regression

Also discuss the impact

of  $f \in \mathcal{F}$  on  $\hat{f}$  and  $\hat{f}^*$ , where  $\hat{f} = \arg \min_{f \in \mathcal{F}} \sum_{i=1}^n (f(x_i) - y_i)^2$  and  $\hat{f}^* = \arg \min_{f \in \mathcal{F}} \sum_{i=1}^n (f(x_i) - y_i)^2 + \lambda \sum_{i=1}^n |f(x_i)|$ .

- Forward Selection
- Backward Selection

as well

as  $f \in \mathcal{F}$  on  $\hat{f}$  and  $\hat{f}^*$ , where  $\hat{f} = \arg \min_{f \in \mathcal{F}} \sum_{i=1}^n (f(x_i) - y_i)^2$  and  $\hat{f}^* = \arg \min_{f \in \mathcal{F}} \sum_{i=1}^n (f(x_i) - y_i)^2 + \lambda \sum_{i=1}^n |f(x_i)|$ .

- Polynomial Interaction (with interaction only)
- Polynomial interaction (power + interaction cases)