

Defense Seminar Opposition

Opponent: Maximilian Pfundstein

Student: Martin Šmelík

June 2020

1 Questions

1. Spatial and Seasonal Information

- (a) As your model does not yet embed spatial information, do you think that this could help and how could you extend your models to embed this information?
- (b) As your model does not yet consider seasonality, do you think that this information would improve the models and how would you modify your model to handle this kind of information?

2. Validity of Results

- (a) As you conclude that all of your models are appropriate for road surface prediction, what exactly is the criteria that leads to this conclusion?
- (b) There seems to be an analysis that has already been conducted using spatial models. Do you know the results and how well they fare compared to your models?

3. Inherent Noise

- (a) You use τ to denote the *Inherent Noise* as well as a precision parameter. You might want to change that.
- (b) Is *Inherent Noise* the same as the mean squared error (MSE)? The definition looks exactly like MSE.

4. Clustering You use clustering to group the data according to different weather conditions.

- (a) As you train your models only within the clusters, don't you think that this limits the capability of your models as they see less data? Would it not be better to take the clustering result as an additional feature?

- (b) Kernel methods could have been used to embed information about closeness of data point which you seem to handle by clustering. Why did you opt for clustering and not something else like kernel methods?
- (c) Did you manually inspect the found clusters? Could you see any patterns or could you identify a cluster representing a specific weather condition?
- (d) If clustering is part of your methodology (pipeline), how do you ensure that the found clusters during training and testing are the same? This refers to the amount of clusters as well as the assigned number embeddings.
- (e) Each method finds a different amount of clusters. Would it not make sense to do the clustering once as a part of your preprocessing and then use the same clusters for all models? Or are they embedded into the models? Please elaborate a bit on this.