To: ENGW3302

CC: Prof. Cecilia Musselman

From: Paul Langton

Subject: Context memo for *When Life Gives You Lemons...* Draft 2

Date: November 5, 2019

When Life Gives You Lemons... is a literature review of a state-of-the-art machine learning technique called gradient boosting. In this memo I will outline the context of gradient boosting and provide key terms which should help the reader understand the intuition behind the method.

The algorithms and mathematics in this review are dense but necessary to convey to scientific peers exactly how gradient boosting methods have advanced. For those readers who are unfamiliar with the field, I do not recommend you spend time reading the math. The concept is not simple and it should not be immediately obvious what makes it so effective, nor why it works *at all*. Reading the Introduction, Conclusion, and Applications, paying close attention to the figures should be enough.

1 Research Context

Gradient boosting is one of many modern machine learning algorithms which take advantage of recent increases in compute capacity to learn a complex function given a large amount of data. At a high level, it works by repeatedly adjusting a set of decision trees to minimize a loss function. The final model is produced by combining the decision trees produced at each step.

It has been the subject of much hype in the community due to its tendency to produce highly accurate and interpretable models for a wide range of problems. Gradient boosting has been referenced widely in literature reviews detailing progress on a specific problem, but currently does not have a review of theoretical developments to the method, which motivates this work.

2 Terminology

- **Decision Tree**: A tree used to model a series of sequential choices, whose nodes represent a decision to be made and leaves represent outcomes or classifications.
- **Learning**: Any process which estimates an unknown function F' given some example input data $\mathbf{x}$ and classifications y by producing a function F.
- **Model/Classifier**: A function F which takes some inputs and produces outputs (classifications) which divide the inputs into individual classes.
- **Gradient (of a function F)**: A vector field whose value at a point P is a vector whose components are the partial derivatives of F at P.
- **Ensemble Learning**: Any learning style which produces a better classifier from a set of weaker classifiers.
- **Online**: A problem setting in which an algorithm may only observe each piece of data one at a time, and only a single time.

- **Sparse**: A description of a piece of data which implies that many of its entries are 0 or missing.
- **Loss Function**: Some function (usually written L) which measures the amount of incorrectness a learned function F achieves on some goal input/output pairs $\mathbf{x}$ and y.
- **Convex Hull (of a set $X \in R^n$)**: The intersection of all convex sets containing X.
- **AUC**: Abbreviation for Area Under Curve, refers to the area under the Receiver Operating Characteristic (ROC) curve, which plots the true positive rate against the false positive rate for a given classifier. A perfect AUC score is 1.