Motivation

Emergency services respond to millions of calls every year throughout the city of San Diego.

- Minutes/seconds in response time difference between life/death
- The ability to effectively predict where emergency incidents will occur could save both lives and money

Our goal is to create a model that can effectively predict where incidents are likely to occur over the next several hours.

Data

- 8 years of emergency incident data from the city of San Diego
- Cleaned, and converted to include latitude and longitude

Temporal-spatial correlation but unknown underlying dynamics.

Problem Descriptions

- Generate grid representation of San Diego
- Problem 1:
  - Predict the number of incidents that will happen per day in each grid cell
  - Helpful for staffing decisions
- Problem 2:
  - Identify which exact areas of San Diego are the most high-risk on a specific day
  - Helpful for emergency anticipation

Models

Decision Tree Regression:
- Used month, day, hour, and grid location as covariates
- Effective, but prone to overfitting
- Non-representative of underlying distribution
- Uses categorical data well

Spatial Clustering:
- Identify locations that are both contiguous and similar in nature based
- Use these clusters in order to build better-informed models
- Steps:
  - Generate a model for each cluster
  - Evaluate and update model on cluster
  - Re-cluster based on model performance

Spatial-Temporal Prospective Excitation Model:
- Utilize temporal and spatial difference in determining likelihood at each locations
- Choose top x% of points that capture Y% of events in a day.

Results

Problem 1:
- Tested our models using 10-fold cross validation
- Experimental Model: RMSE:
  - Decision Tree Regression 0.7010 (incidents/hour)
  - Decision Tree Regression with Max Depth 10 0.4827 (incidents/hour)
  - Spatial Clustering with linear regression 0.4583 (incidents/hour)

Problem 2:
- Tested our model on ten randomly selected days
- Trained model on the weeks leading up to each of those days
- Model selects the top 1% of at-risk locations

- Small variation and little information returned by solutions to problem 1 leads to the need for a solution to problem 2
- Temporal fluctuations are important, but difficult to anticipate
- Our solution to problem 2 would effectively inform a dynamic resource allocation model

Next Steps:
- Model each type of incident separately
- Predict type of incident as well

References