6.S097 Final Presentation
Moin Nadeem, Ramya Durvasula, Faraaz Nadeem, Amy Fang
Problem

- We want to be able to guess house prices
- Most house price estimates do not incorporate visual aspects, but there is obviously some correlation between how a house looks and its cost
- This problem hasn’t been previously solved: no one uses image in price prediction
Our Solution

- Only used the raw pixel data of houses, no data about square footage, number of beds/baths, last sale price, relative location, etc
- Augmented the training data to prevent overfitting (adding pictures with distortion, rotation, translation, reflection, etc.)
- 8 “buckets” of price ranges
train_datagen = ImageDataGenerator(
    rescale=1./255,
    shear_range=0.2,
    zoom_range=0.2,
    horizontal_flip=True)
Model Architecture

- Batch Size = 32, 16
- 32x32 Filter Size
- 2x2 Max Pooling
- Buckets:
  - 0 to 100k
  - 100k to 200k
  - 200k to 350k
  - 350k to 550k
  - 550k to 800k
  - 800k-1.1m
  - 1.1-1.45m
  - 1.45m-3m
Visualization

(all convolution filters available at http://moinnadeem.com/housing_prediction/conv_1/)
Visualization… oh!

(all convolution filters available at http://moinnadeem.com/housing_prediction/conv_1/)
Visualization, done properly.

(all convolution filters available at http://moinnadeem.com/housing_prediction/conv_1/)
Visualization

**model accuracy**

- Train
- Test

**model loss**

- Train
- Test
Performance

- Model performed at 40% accuracy on the testing set
- Given that we had 8 buckets, the accuracy of a random model would have been 12.5%, so this is a decent sized improvement

- Improvements
  - Adam optimizer over Adadelta / RMSprop
    - “Momentum” in Adam proved to be pivotal
  - Changed batch size from 32 to 16
    - Also solved memory issues
  - Used an ImageDataGenerator to skew in order to prevent overfitting.

- We may be able to substantially improve the accuracy if we added more data (currently at 3k)
Learned Lessons

- Data is everything
- Use more epochs
- What if we added zip codes or did parameter tuning?

"Today we learned how to count to one."
Future Improvements

- Add relative location features such as zipcode, lat/long values
  - Merge layer
- Add overhead view, indoor view, other standard features such as number of bed/baths, square footage, etc.