Pixelating Geo-Diversity

David Rodriguez, Jingchuan Chen, Dhia Mahjoub
Before we start:

https://github.com/DavidRdgz/hitbsec-notebook
Why us:

- We are maintaining a domain risk-score API used by our customers
- Last year we’ve been deploying tensorflow models into production
- GLMs
- Convolutional Neural Networks
- 500 Jobs a day in Complex Workflows (Hadoop Based)
Section 1
Introduce Requester Geo-Popularity Data

Section 2
Introduce Exponential Moving Averages (EMAs)

Section 3
Convolutional Neural Networks
Section 1

Introduce Requester Geo-Popularity Data

+ Intuitions for modeling requests
+ Modeling requests at scale
Countries
+
Requests
Carding sites
Altcoins sites

US, IT, GB, CA, UA, NG, RU
Countries
+
Counts
Counts
+
Time
Section 2

Introduce Exponential Moving Averages (EMAs)
The Max
Exponential Moving Average

\[ S_t = \alpha Y_t + (1 - \alpha)S_{t-1} \]

Optional

\[ S_t = \alpha[Y_{t-1} + (1 - \alpha)Y_{t-2} + (1 - \alpha)^2Y_{t-3} + \cdots] \]
Convolutional Networks
Section 2

Convolutional Neural Networks

+ Why these models?
+ Inputs and Layers

Input Layer
Example of non-convolution net

Input Layer

Hidden Layer 1

Hidden Layer 2
Convolution

Input Layer

4 by 4
Input Layer

Convolution

Stride by 2

Input Layer
Convolution Output

Sort of 4 outputs
Max by Pool

The max

2 by 2 Pool
Fully Connected Layer
Conclusion

- Introduce Requester Geo-Popularity Data
- Introduce Exponential Moving Averages (EMAs)
- Convolutional Neural Networks on Geo-Popularity
Questions?