

Machine Learning 101

PyCon 2015

Kyle Kastner

LISA / MILA

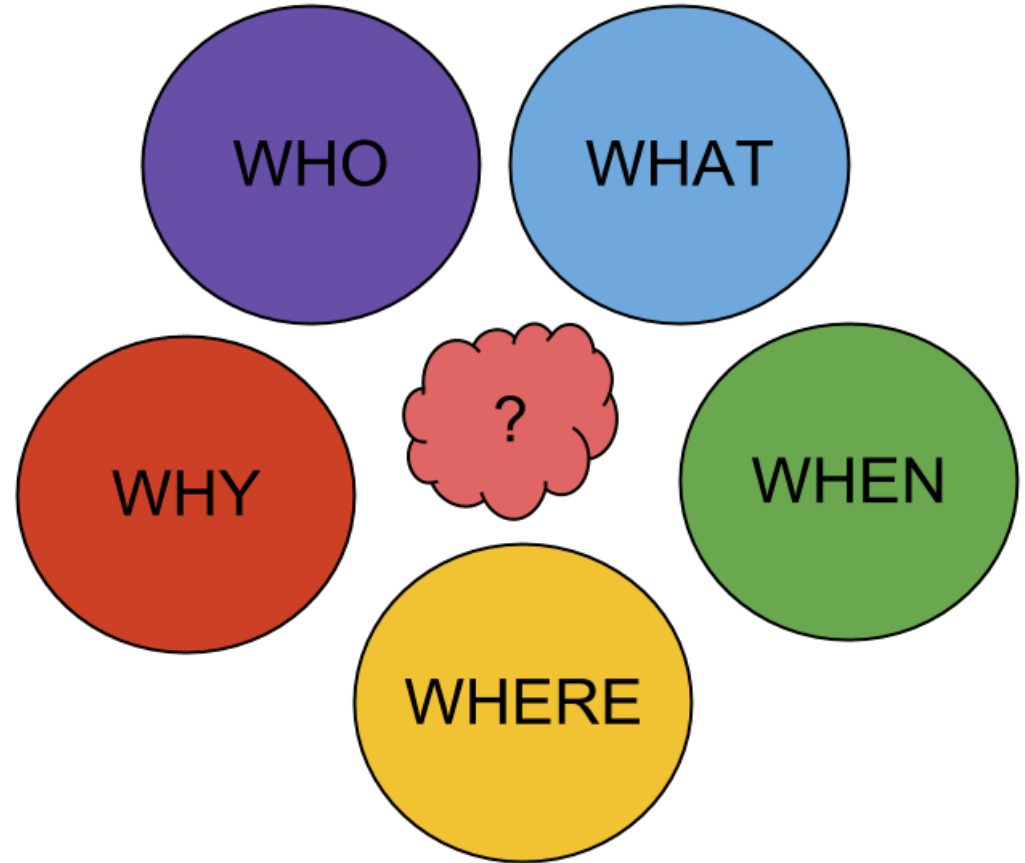
Université de Montréal

Follow along!

<https://github.com/kastnerkyle/PyCon2015>

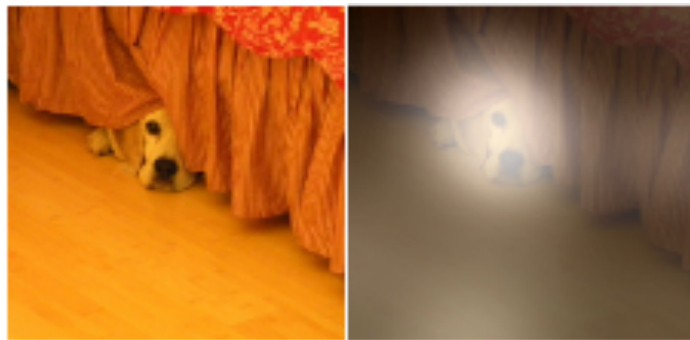
What is Machine Learning?

- Automation
- Data Analysis

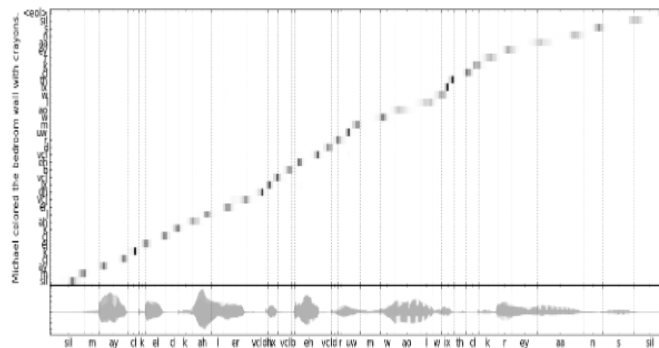


Applications

- Speech processing
 - Speech to text, text to speech
- Image processing
 - Self driving cars
- Natural Language Processing
 - Automatic translation
- Advertising
 - Click Through Rate (CTR) (talk @ 12!)
- Recommendations
 - Amazon, Yelp, Netflix...



A dog is standing on a hardwood floor.



Automation Spectrum



Handcrafted Rules

- if elif elif elif
- **DON'T TOUCH** code
- Magic constants

Statistics

- linear models
- p values
- Bayesian stats
- MCMC sampling

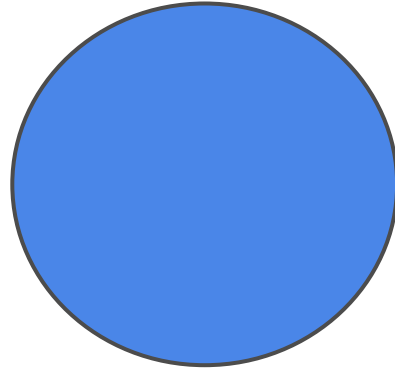
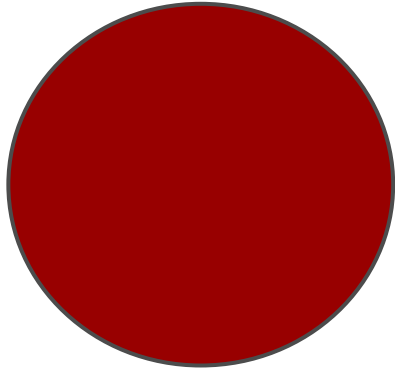
Machine Learning

- K-means
- SVM
- Random Forests

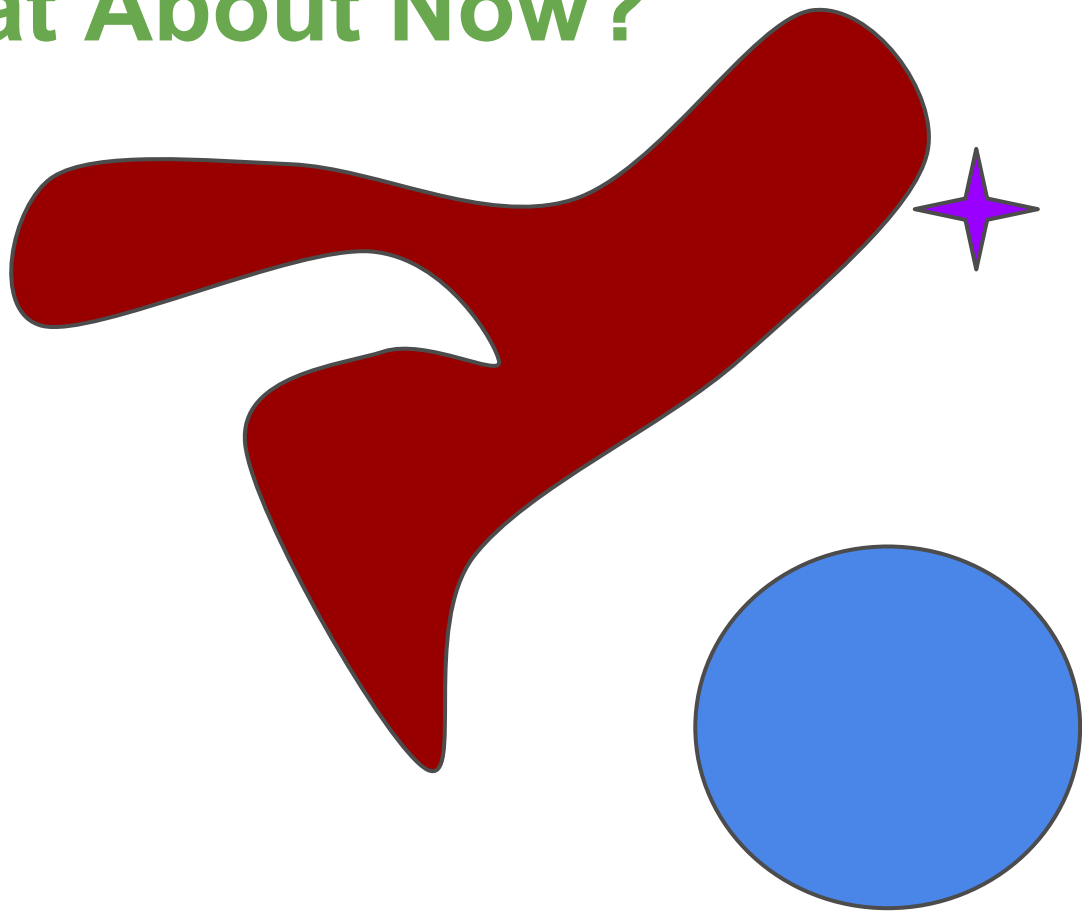
Deep Learning

- Neural networks
- Autoencoders
- Recurrent net
- Convolutional net

A Test



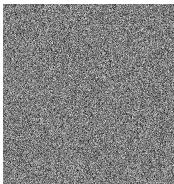
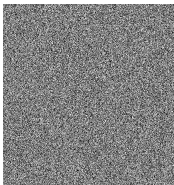
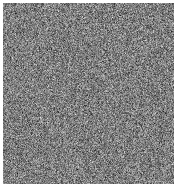
What About Now?



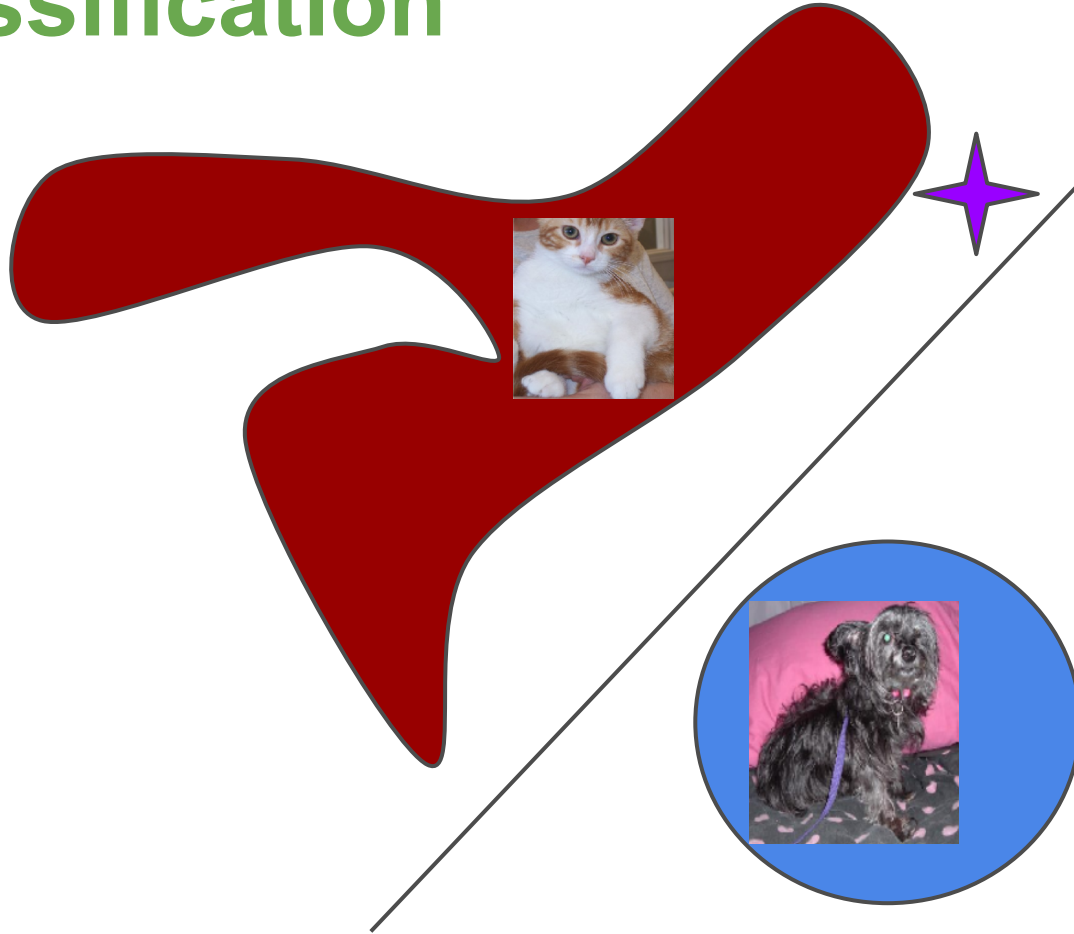
ifold Hypothesis



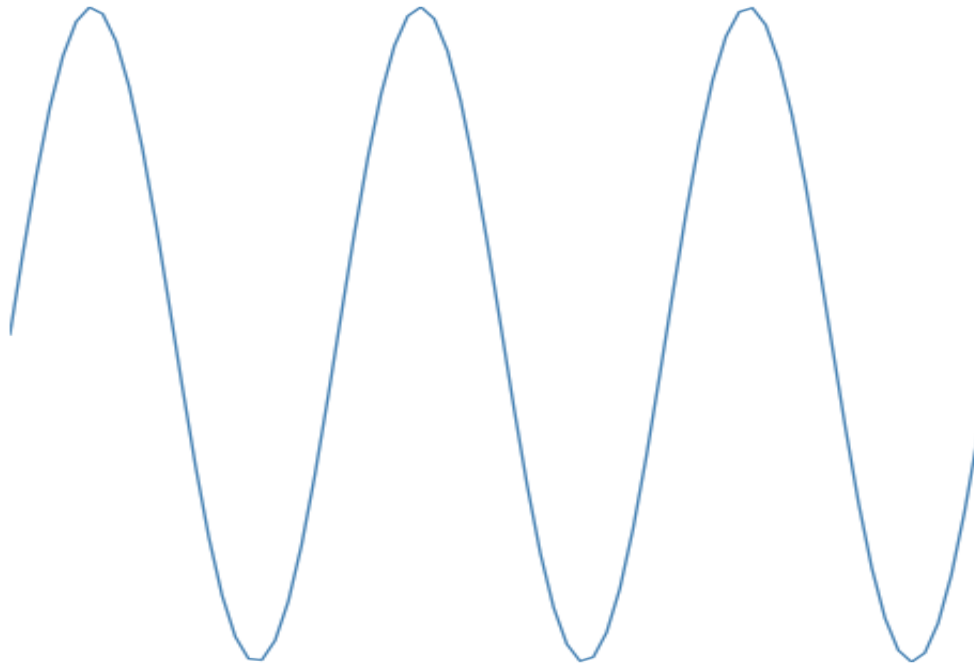
A large red swastika symbol with a black outline, centered on a white background. To the right of the symbol is a purple four-pointed star. Below the star is a small inset image of a ginger and white kitten sitting next to two colorful Easter eggs. In the bottom right corner, there is a blue circle containing a black and white photo of a small, scruffy dog. In the bottom left corner, there is a small square of gray static noise.



Classification

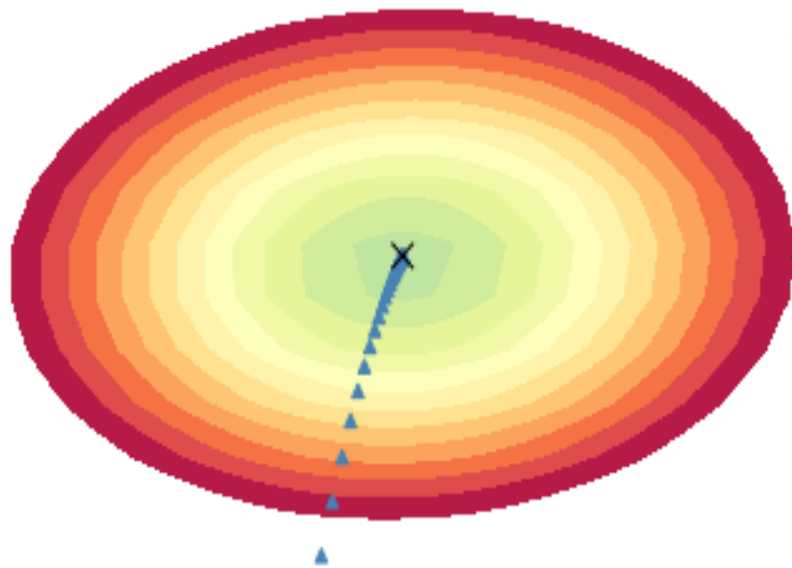


Regression



Learning Functions

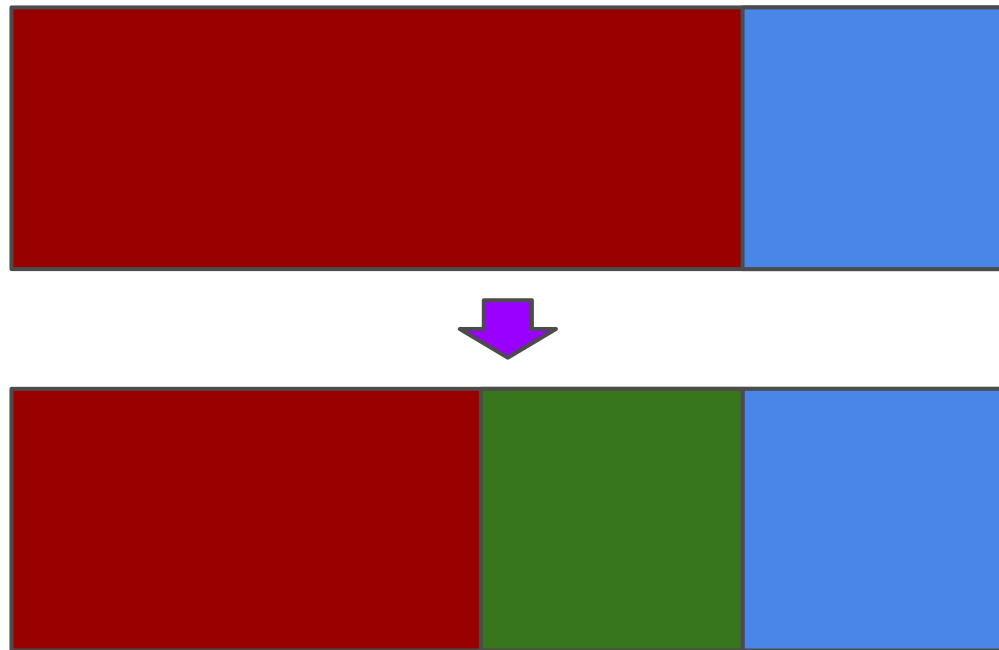
```
def recommend_movies(seen_movies, ratings):  
    ...  
    return best_unseen_movies  
  
def french_to_english(french_text):  
    ...  
    return english_translation  
  
def what_is_in_this_image(image):  
    ...  
    return text
```



$$y = f(x); y = f(x|\theta); p(x|\theta) \propto p(\theta|x)p(x) \quad (\text{Bayes Rule})$$

Train/Valid/Test

- Split current data
- Evaluate
- Typical split
 - 80% training
 - 20% validation



- Testing data answers unknown
- Want systems to work on new data!
- This approach *simulates* new data

What should I use?

- I recommend one of two packages
 - Anaconda, from *Continuum.io*
 - Canopy, from *Enthought*
- Both excellent!

Anaconda: <https://store.continuum.io/cshop/anaconda/>

Enthought: <https://store.enthought.com/>



Examples

List of Resources

- Google Python Class <https://developers.google.com/edu/python/?csw=1>
- Numpy tutorial http://wiki.scipy.org/Tentative_NumPy_Tutorial
- Numpy to Matlab table http://wiki.scipy.org/Tentative_NumPy_Tutorial
- scikit-learn documentation <http://scikit-learn.org/stable/tutorial/index.html>
- scikit-learn tutorial slides https://github.com/ogrisel/parallel_ml_tutorial
- more tutorial slides https://github.com/jakevdp/sklearn_pycon2015/
- Coursera ML course (octave/Matlab) <https://www.coursera.org/learn/machine-learning>
- Stanford UFLDL http://ufldl.stanford.edu/wiki/index.php/UFLDL_Tutorial
- Ian Goodfellow's Intro to Theano https://github.com/goodfeli/theano_exercises
- Theano notebooks <http://nbviewer.ipynb.org/github/jaberg/IPythonTheanoTutorials/tree/master/ipynb/>
- Theano Deep Learning Tutorial <http://deeplearning.net/tutorial/>
- Machine Learning for Vision http://www.iro.umontreal.ca/~memisevr/teaching/ift6268_2015/index.html
- Representation Learning <https://ift6266h15.wordpress.com/>
- Coursera NN course <https://www.coursera.org/course/neuralnets>

<https://github.com/kastnerkyle/PyCon2015>



@kastnerkyle



@kastnerkyle

Thank You!

References

[1] Taken from Wikipedia

http://en.wikipedia.org/wiki/File:EM_Spectrum_Properties_edit.svg

[2] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhutdinov, R. Zemel, Y. Bengio.

Show, Attend and Tell: Neural Image Caption Generation with Visual Attention

<http://arxiv.org/abs/1502.03044>

[3] J. Chorowski, D. Bahdanau, K. Cho, Y. Bengio.

End-to-end Continuous Speech Recognition using Attention-based Recurrent Neural Networks

<http://arxiv.org/abs/1412.1602>

[4] J. Elson, J. Douceur, J. Howell, J. Saul. [Asirra: A CAPTCHA that Exploits Interest-Aligned Manual Image Categorization](#). In Proceedings of 14th ACM Conference on Computer and Communications Security (CCS), Association for Computing Machinery, Inc., Oct. 2007

[5] G. Hinton, P. Dayan, M. Revow. Modelling the Manifolds of Images of Handwritten Digits.

<http://www.cs.toronto.edu/~fritz/absps/manifold.pdf>

[6] Bayes Rule. http://www.eecs.qmul.ac.uk/~norman/BBNs/Bayes_rule.htm