

An Engineer's Trek into Machine Learning

Satish Chandra Gupta
Co-founder, Slang Labs

A Typical Trek



- Irresistible Allure
 - Is the trail worth it?
- Study the Terrain
 - Key features
 - Difficulties and dangers
- Train Hard
 - Plan, prepare
- Check your Gears
 - Map and Compass
- Adventure Begins



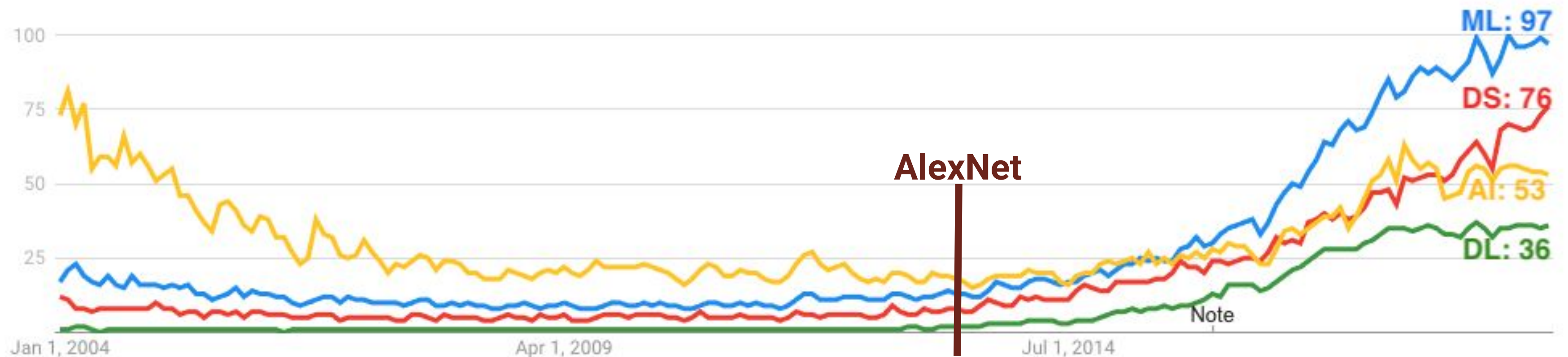
Irresistible Allure

—

In God We Trust; Others Must Provide Data.



Google Search Trends:



AI: Artificial Intelligence, DS: Data Science, ML: Machine Learning, DL: Deep Learning

ML/AI is all around us



Consumers

- Web Search: Result ranking
- Mail: Spam Detection, Inbox classification, Sentence completion
- Maps: ETA
- News: Story clustering
- Web/App Ads: Google, Facebook
- Recommendation Systems: Amazon, Netflix, Facebook, Quora
- Personal Assistants: Siri, Alexa, Google

Enterprises

- Fraud detection: Banks, Amazon
- Pricing/Financial Modeling
- Customer churn prediction
- Equipment failure prediction
- Network intrusion detection
- Customer segmentation
- Sentiment analysis
- Image/video analysis
- Speech/audio analysis

Study the Terrain

Artificial Intelligence vs. Machine Learning vs. Data Science

scgupta.link/ml-intro-for-developers



Artificial Intelligence (AI)

Intelligence demonstrated by machines by performing like humans.

Machine Learning (ML)

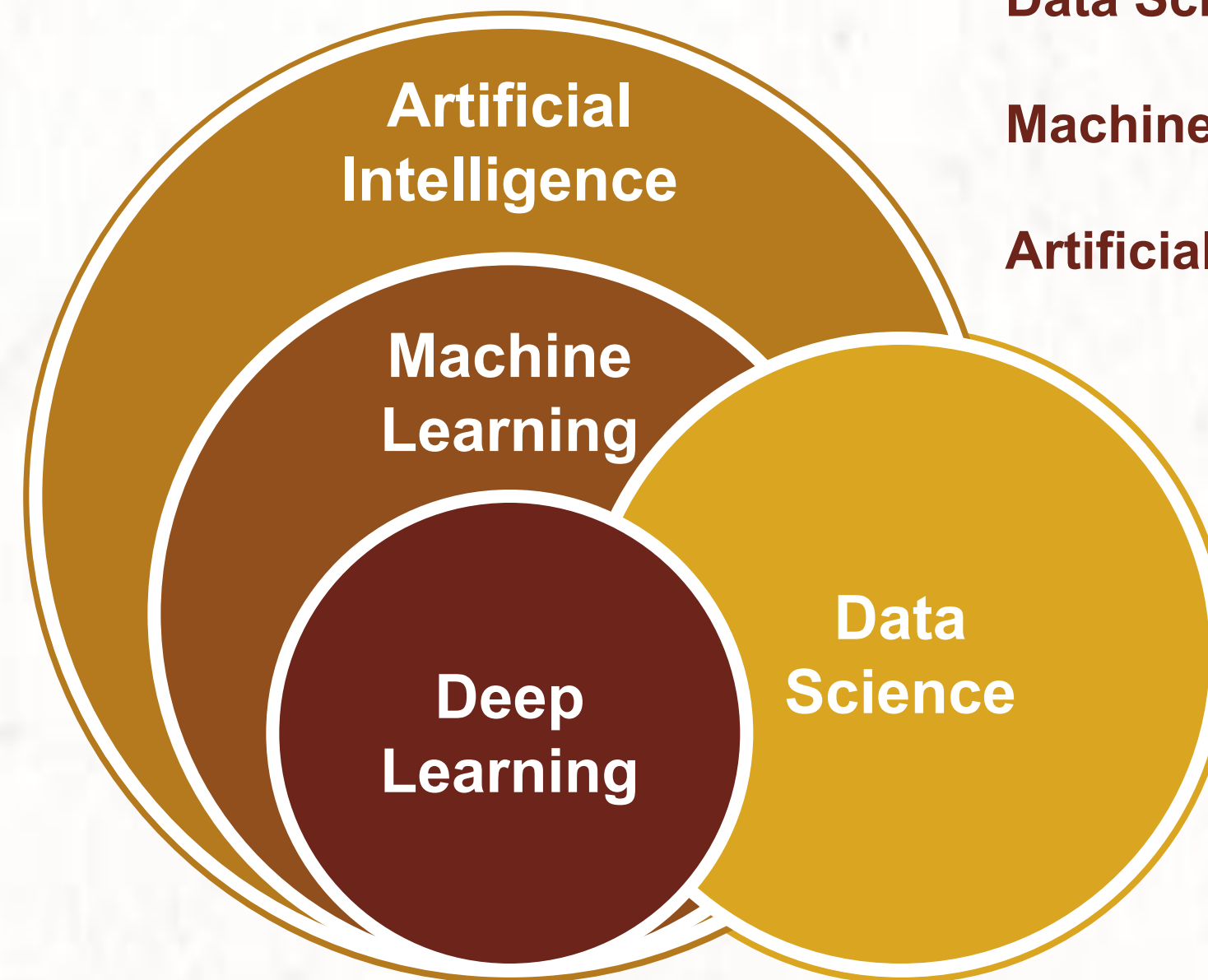
Making predictions with statistical models, by recognizing patterns in the data, without requiring explicit instructions.

Deep Learning (DL)

Machine Learning methods using deep neural networks.

Data Science (DS)

Use of statistical techniques to manually extract insights from data.






Data Science produces insights.

Machine Learning produces **predictions**.

Artificial Intelligence produces **actions**.

Deep Learning is used on:

-  Images, videos
-  Natural language text
-  Speech audio

Statistics
Linear Algebra
Calculus

Mathematics

Programming

Data Engineering
Software Engineering
MLOps

© Satish Chandra Gupta



CC BY-NC-ND 4.0 International Licence
creativecommons.org/licenses/by-nc-nd/4.0/

scgupta.me 

twitter.com/scgupta 

linkedin.com/in/scgupta 

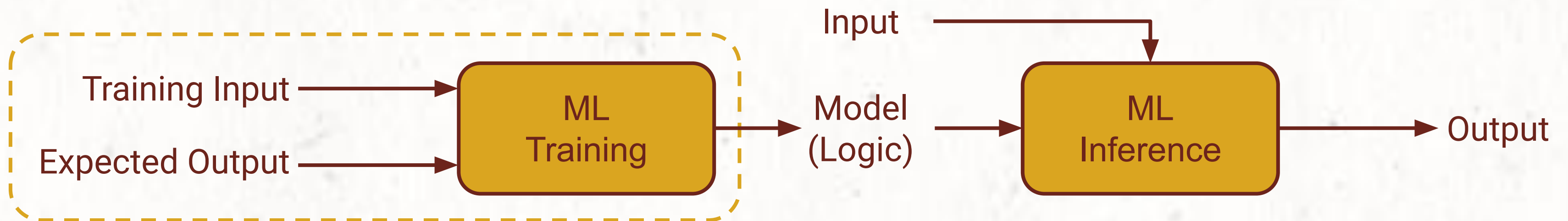
Traditional Program vs. Machine Learning



Traditional Programs: Define algo/logic to compute output



Machine Learning: Learn model/logic from data



Example: Email Spam Detector



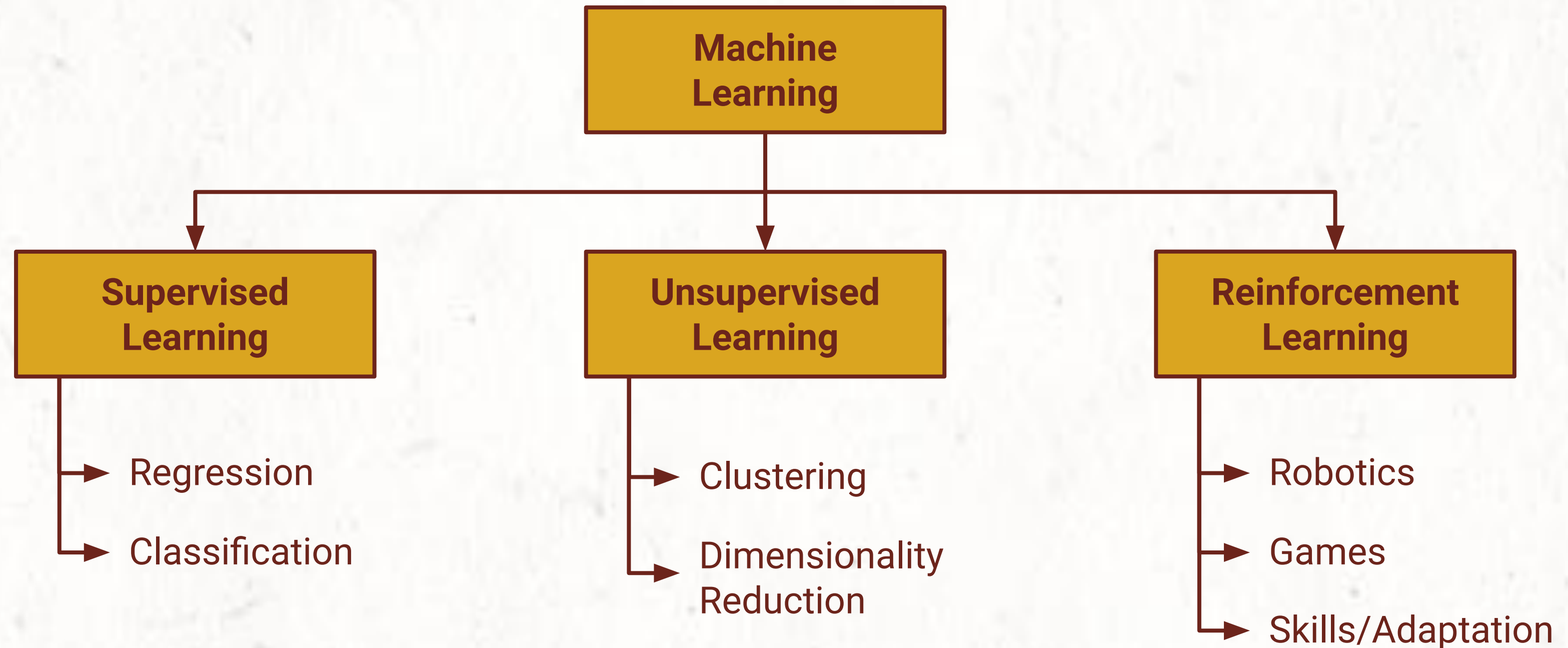
Traditional Programming

- How do human identify spams?
- Enumerate as exhaustive list of rules:
 - Word **FREE** occurs several times
 - Phrases like: Weight Loss
 - You have won lottery
 - ...
- As spammer changes tactics
=> continuously update these rules
- Knowledge / Expert System AI

Machine Learning

- Prepare Data Set: human have tagged mails to Spam / No spam
(Essentially user's Spam folder)
- Train / test / tune models, select best
- Use it to move mails to Spam
- Track: if user moves a mail from Inbox to Spam, or vice versa; add these to Data Set
- Remain up-to-date with new spam trends

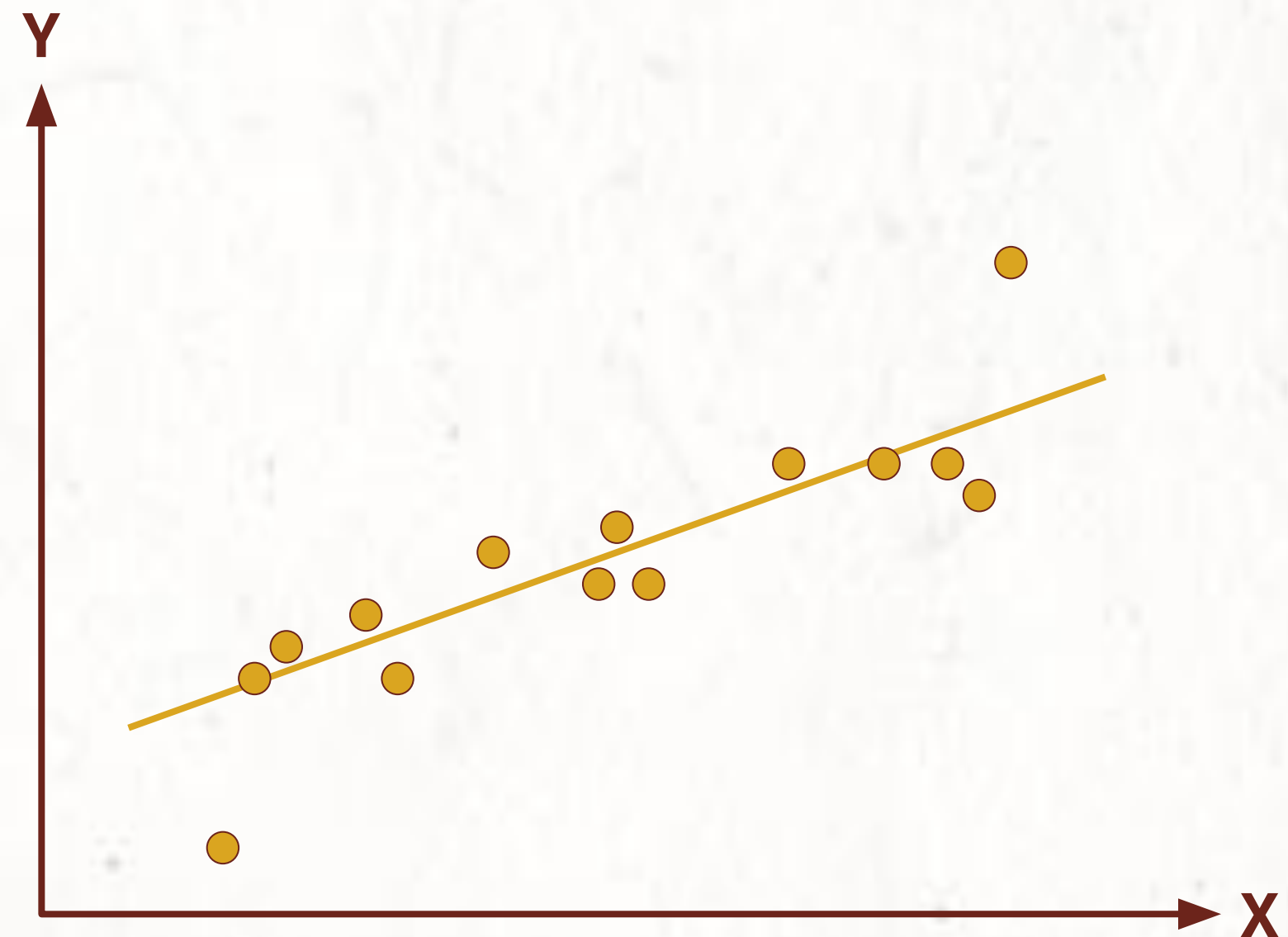
Machine Learning



Regression



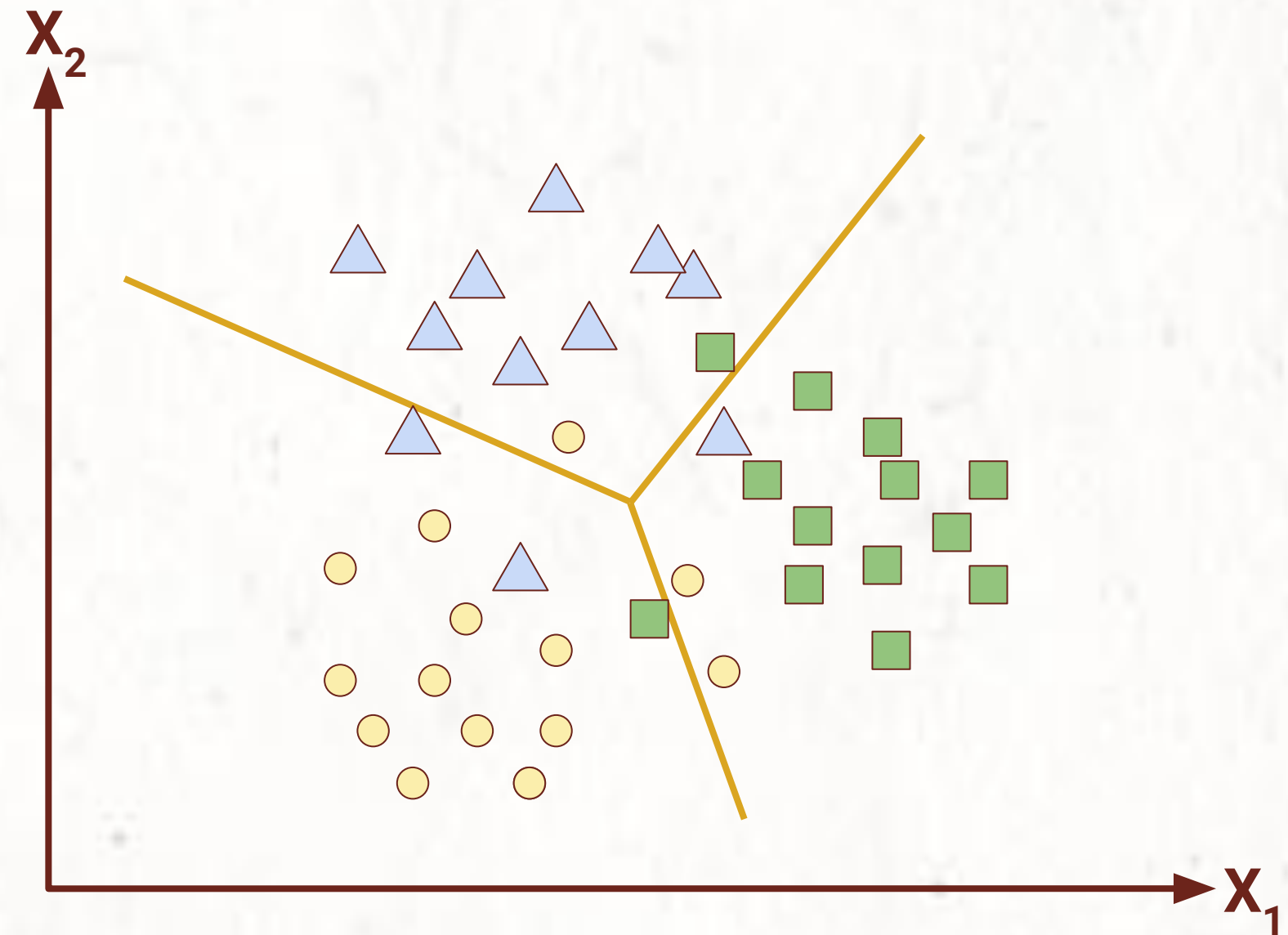
- Estimating value of dependent variable with one or more independent variables
- $Y = f(X)$
Y: outcome
X: features
- Estimate value of house, given
 - Size
 - Location
 - Number of Bedrooms
 - Number of Bathrooms



Classification



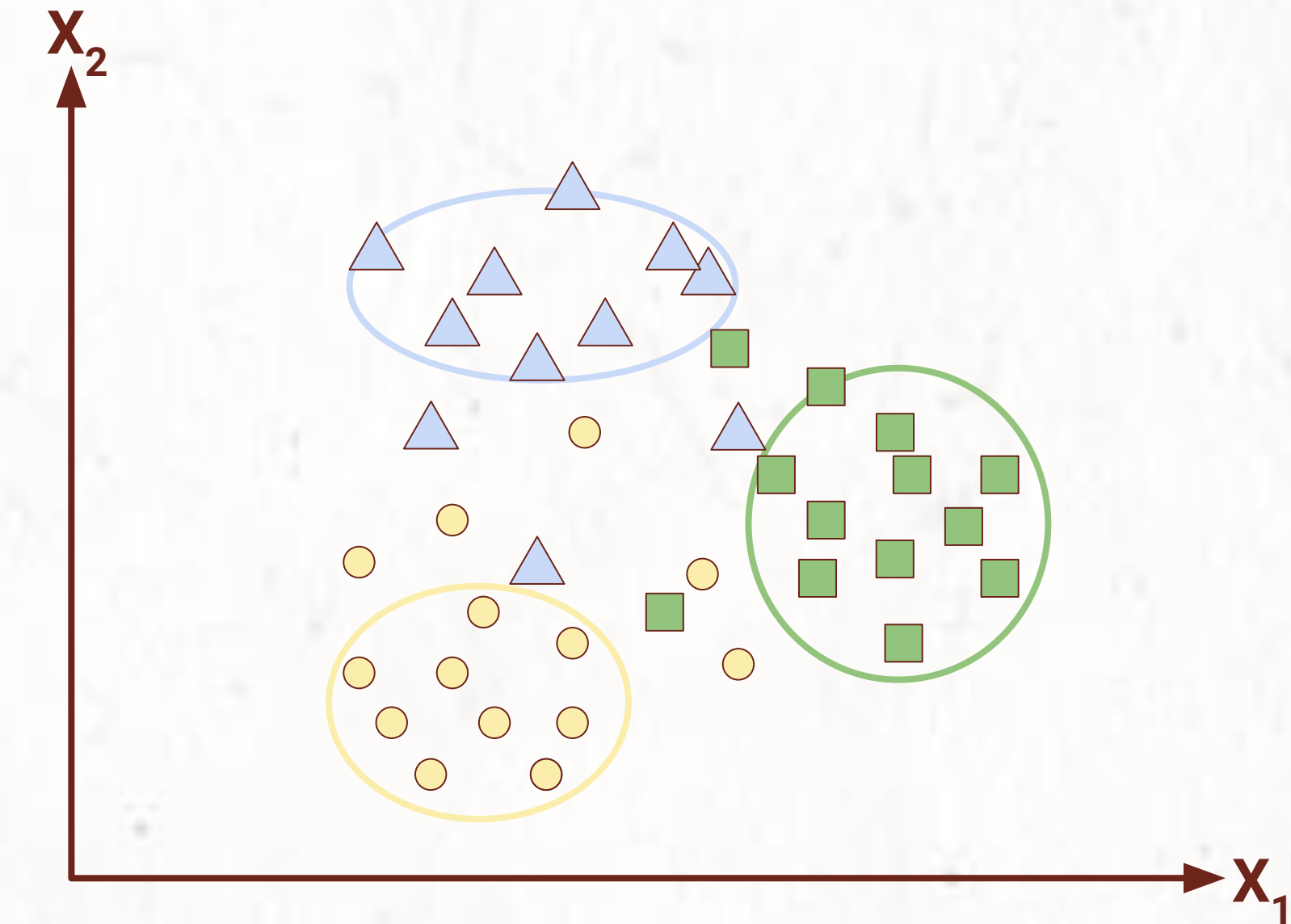
- Identifying class/group from characteristics of an object
- $Y = f(X)$
Y: labels, finite set
X: features
- Given a photo, identify whether it is a
 - Car
 - Truck
 - Motorcycle



Clustering



- Group objects into clusters of similar objects
- Unsupervised: Number and definition of clusters may not be known in advance
- Some notion of similarity
 - Connectivity, Centroid, Density, etc.
- Cluster similar news articles
(based on topics that are not predefined)



No magic bullet: It is all in the data



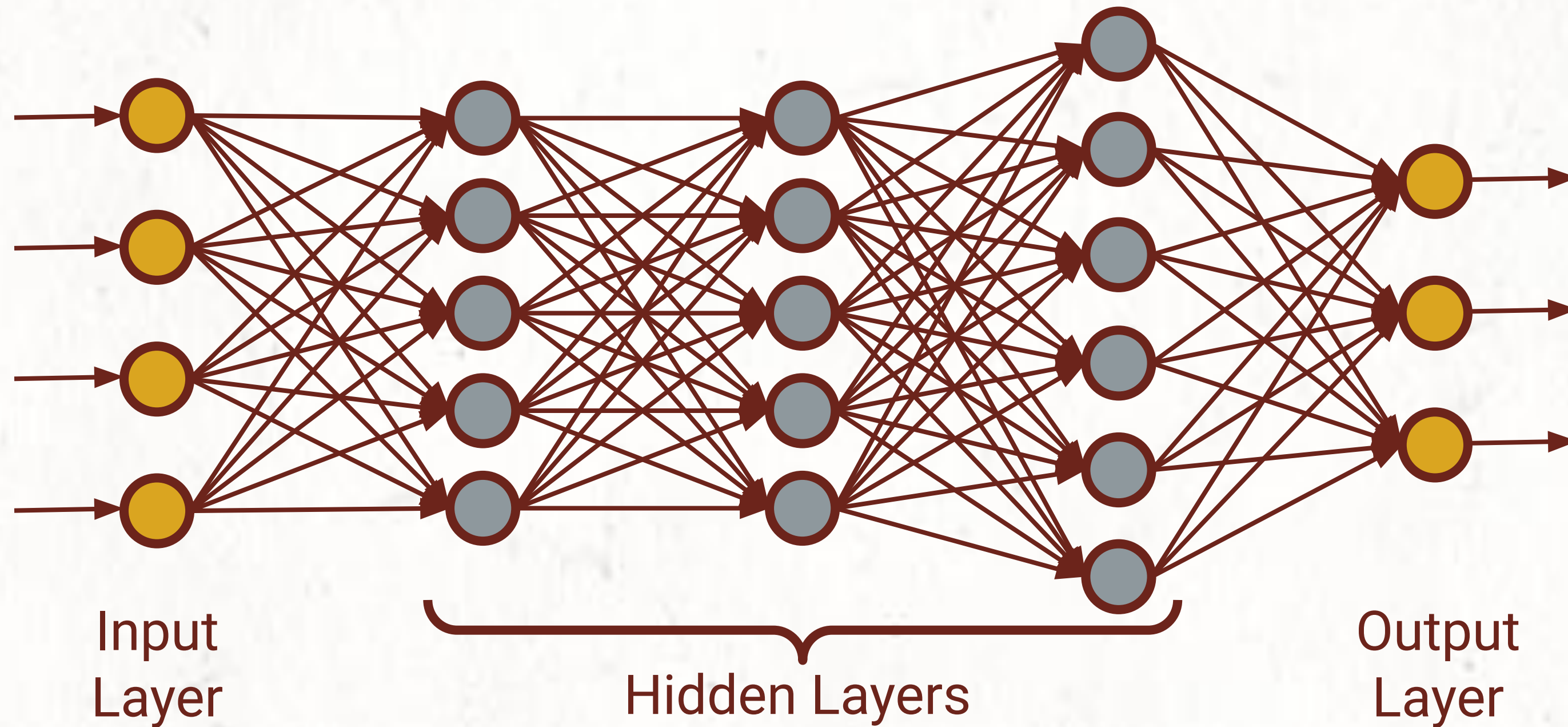
- Data is logic
 - Garbage In \Rightarrow Garbage Out
- Train only for Truck and Car, model can not
 - Detect motorcycle
 - Distinguish Blue/Red Car and Blue/Red Truck
- Important to build intuition
 - Data Visualisation
 - How human will do what you are doing with ML

Deep Learning

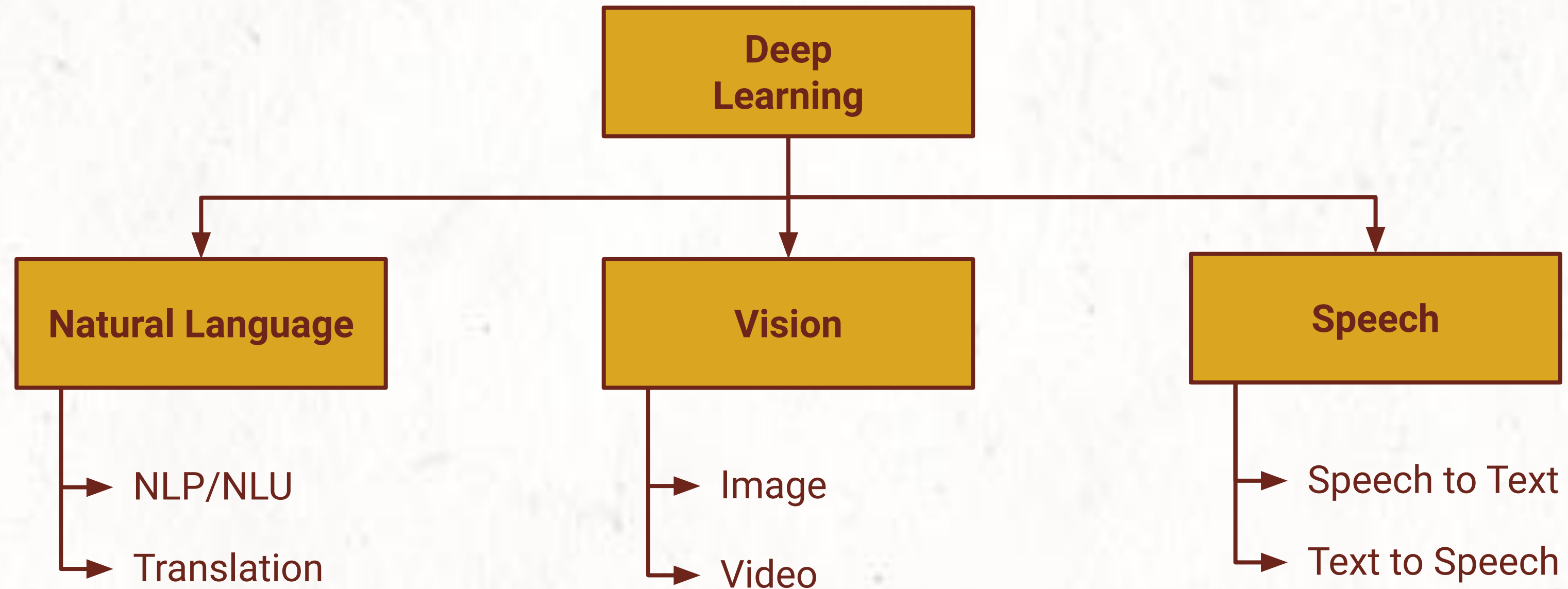


- Deep Neural Network (DNN)
 - Convolutional CNN), Recurrent (RNN), Long Short-Term Memory (LSTM), ...
- Problems are same
 - Regression, Classification, ...
- Type of data: Structured vs. Unstructured
 - DNN used mostly on unstructured data
- Unstructured Data
 - Natural Language, Vision, Speech

Deep Neural Network



Deep Learning



Tools & Frameworks



Data Science / Machine Learning

- SciKit-Learn
- Pandas
- NumPy
- SciPy

- Matplotlib
- Seaborn
- Jupyter Notebook

Deep Learning

- TensorFlow
- PyTorch
- Keras

Cloud Deployment

- Google Cloud AutoML
- Amazon SageMaker
- Azure ML

Train Hard

—

Online Courses & Tutorials



- Machine Learning Basics
 - Andrew Ng's famous course: <https://www.coursera.org/learn/machine-learning>
- Machine Learning for practitioners with Python
 - <https://www.udemy.com/python-for-data-science-and-machine-learning-bootcamp/>
- TensorFlow
 - <https://www.tensorflow.org/tutorials>
- Deep Learning
 - <https://www.deeplearning.ai/>
- Cloud: Google, Amazon, Microsoft have online tutorials for their platforms
- Kaggle:
 - <https://www.kaggle.com/>

Books



- The Elements of Statistical Learning: Hastie, Tibshirani, Friedman
 - <https://web.stanford.edu/~hastie/ElemStatLearn/>
- An Introduction to Statistical Learning: James, Witten, Hastie, Tibshirani
 - <http://faculty.marshall.usc.edu/gareth-james/ISL/>
- Deep Learning: Goodfellow, Bengio, Courville, Bach
 - <https://www.deeplearningbook.org/>
- Pattern Recognition and Machine Learning: Bishop
 - <https://www.microsoft.com/en-us/research/people/cmbishop/#!prml-book>
- Machine Learning Yearning: Andrew Ng
 - <https://www.deeplearning.ai/machine-learning-yearning/>

Check your Gears

—

Data Scientists vs. Software Engineers



Data Scientists

- Strong math, especially statistics
- Expert in playing with data
- Design efficient models

Potential Weaknesses:

- May not appreciate need of continuous testing for correctness guarantees
- Typically built batch systems, may require significant (re)work to take it to production

Software Engineers

- Strong programming skill
- Expert in highly scalable applications
- Design/test/deploy 24x7 running system

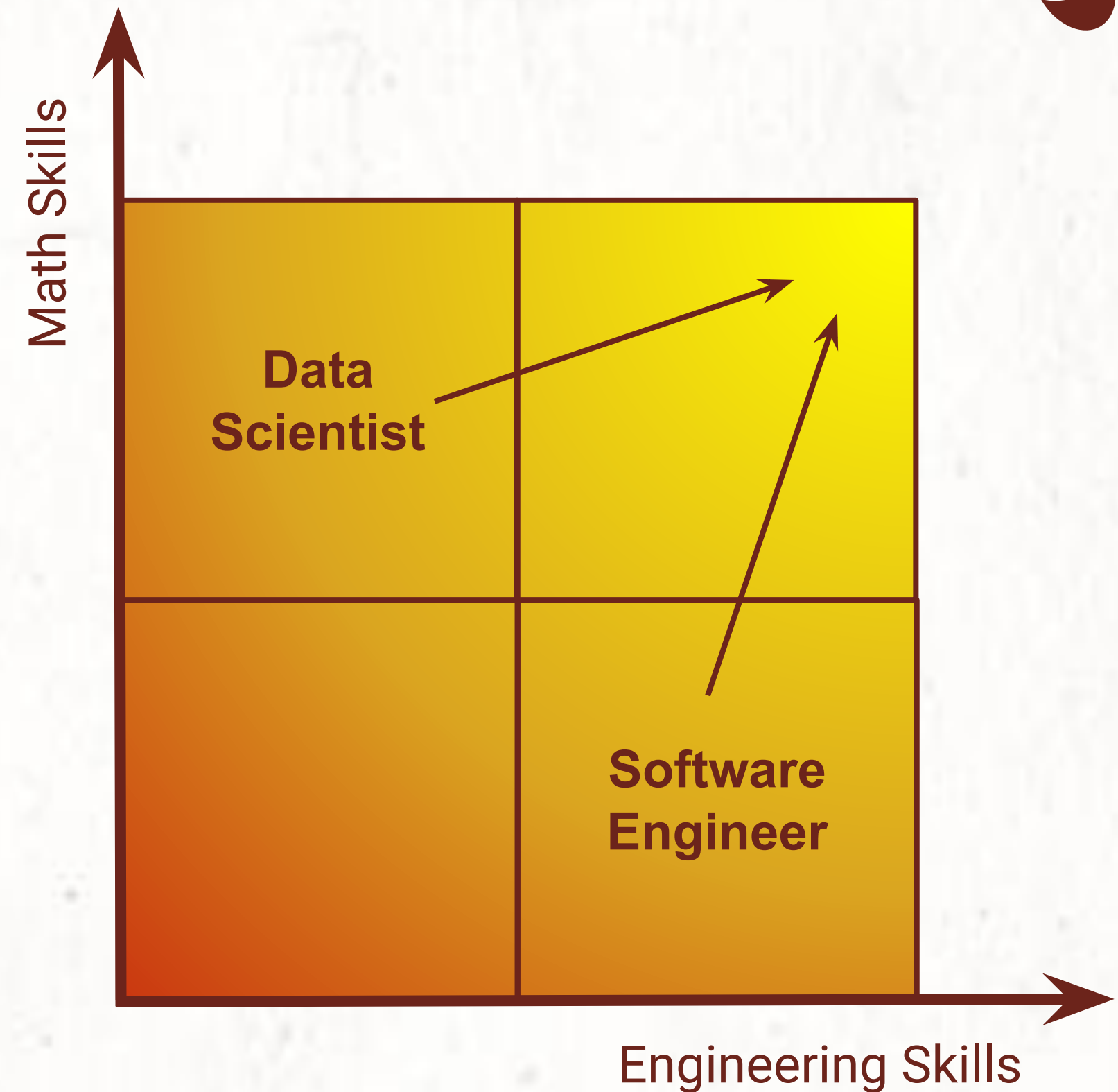
Potential Weaknesses:

- Deterministic logic is ingrained. Struggle with notion of statistical correctness (how a test can fail if no code has changed?)
- if weak in math, model is a black box

Map & Compass



Miles to go before I sleep...



Tips for Operationalizing ML/AI



- Production can be graveyard of un-operationalized ML
- Figure out Data Strategy: Quality, Schema & Catalog
- Be seduced by impact, and not by cool tech
- Do not create ML models in silos
- Consider the scale of data, and feedback loop
- Be mindful of deployment and production environment
- Simplicity often trumps cleverness in production

Adventure Begins



Key Takeaways



- AI is real, future is here.
- “AI is the new electricity.” - Andrew Ng
- But it is the just the beginning of AI adventure.
- You can learn anything by training.

Best of Luck, Bon Voyage !!!

**MAY
THE
FORCE
BE WITH
YOU**



Thank You!

scgupta.link/ml-intro-for-developers

 twitter.com/scgupta

 linkedin.com/in/scgupta

 scgupta.me