



MANGO  
SOLUTIONS

# What Does a Statistician Need to Know about Machine Learning?

Aimee Gott

Senior Consultant

✉ [agott@mango-solutions.com](mailto:agott@mango-solutions.com)

🐦 [@aimeegott\\_R](https://twitter.com/aimeegott_R)



# My Story...



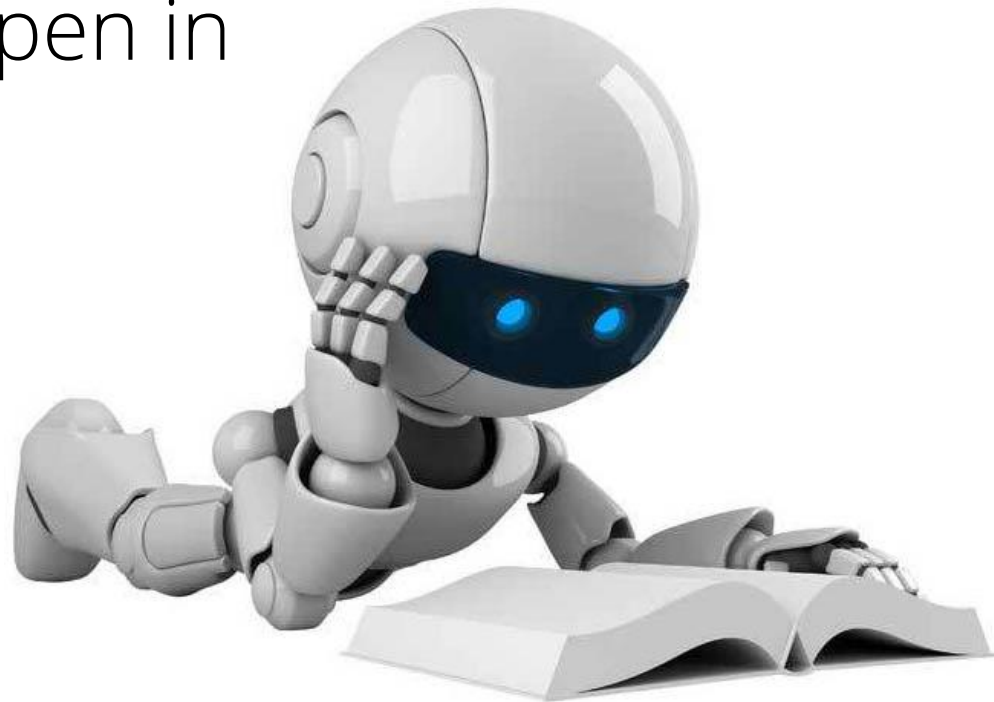
# Agenda

- What is machine learning?
- The language of machine learning
- Models, models and more models



# What is Machine Learning?

Learning from what  
happened in the past to  
predict what will happen in  
the future



# What about Statistics?

Learning from what has happened in the past to explain why it happened



# Machine Learning in Action

anagram



All

Images

Shopping

News

Videos

More

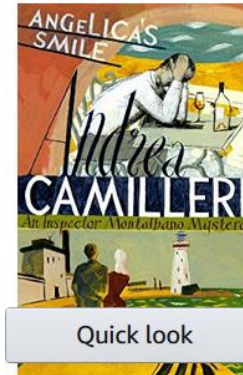
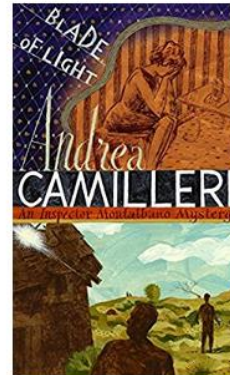
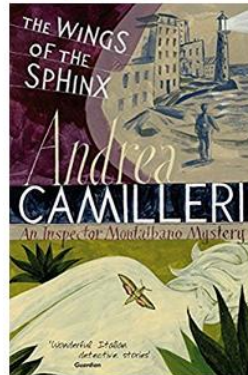
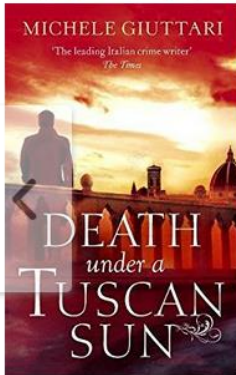
Settings

Tools

About 18,400,000 results (0.70 seconds)

Did you mean: **nag a ram**

Inspired by your shopping trends



Quick look



# What is Machine Learning?

- Using data to identify patterns
- Predicting new events
- Learning from the new events to update the model



# Is ML Really Something Different?

Yes...  
....and no!





# The Language of ML

- Target
- Features
- Supervised learning
- Unsupervised learning
- Ensemble methods
- Feature Engineering
- Bagging
- Boosting



# Supervised Learning

- Problems for historic data that includes the outcome
  - Regression
  - Decision Trees
  - Neural Networks (Deep learning)
  - Support Vector Machines



# Unsupervised Learning

- Problems with no known outcome
  - Clustering
  - Dimensionality Reduction (PCA)



# Ensemble Methods

- Combining approaches to build better models
  - Random Forest
  - Gradient Boosting



# So Many Models!!!

- Statisticians typically focus on one
- Data Scientists are familiar with many



# Building a Model

- Test vs Train
- Cross Validation
- Feature Engineering



# Feature Engineering

- One-hot-encoding
- Discretization



# What Do You Need to Know?

- Machine Learning is not completely new
- Different words, same meaning
- Overview of many models, less focus on one



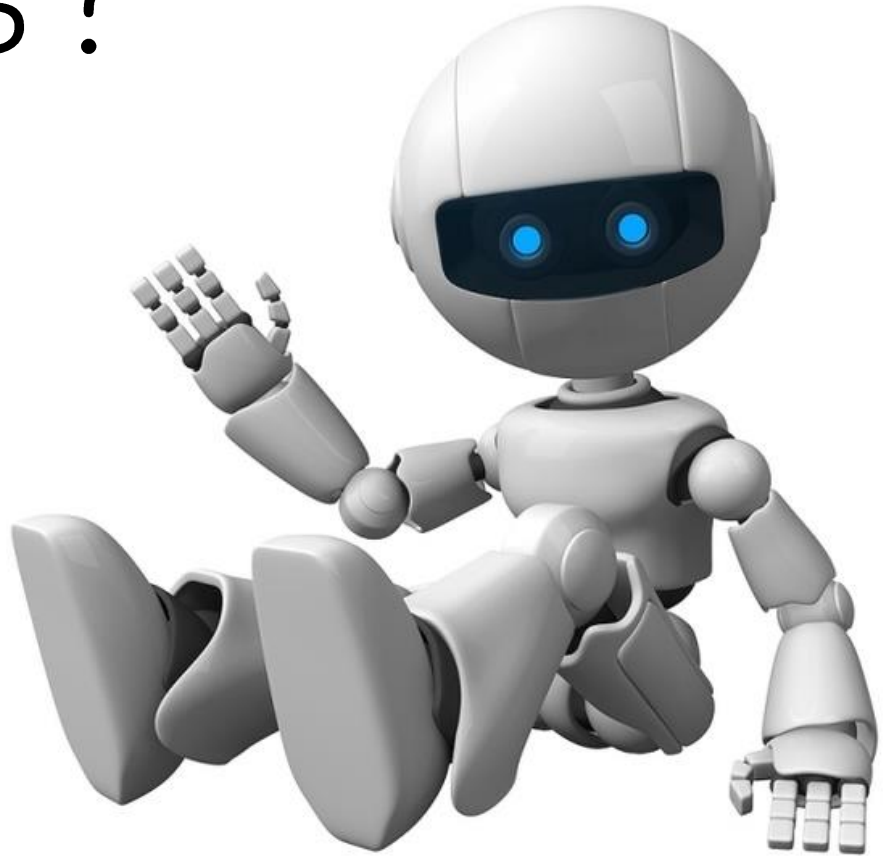


# Want to Know More?

- Introduction to Machine Learning in R
  - October 2017, RSS



# Questions?



Aimee Gott

✉ [agott@mango-solutions.com](mailto:agott@mango-solutions.com)