Applied Data Science

Day 1 - Course Welcome

@cambridgespark

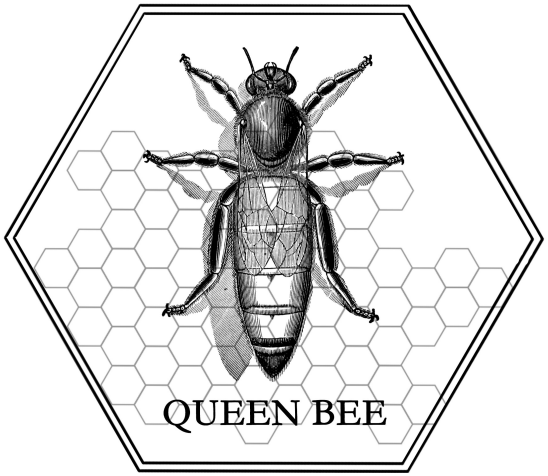
CAMBRIDGE SPARK

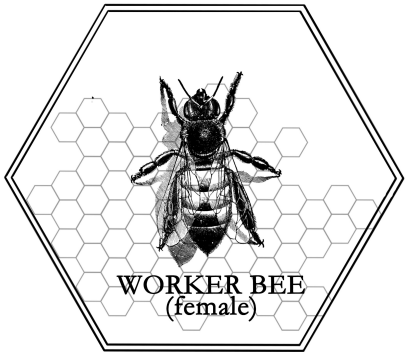
Module outline

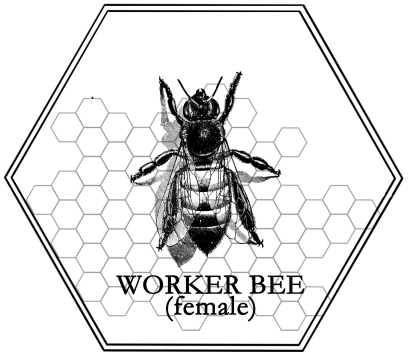
- Console tools:
  - `git`, `conda`, and `jupyter`
- Python packages:
  - `numpy`, and `pandas`
- Mathematics for Data Scientists:
  - Linear Algebra
  - Probability
  - Optimisation
- **K.A.T.E.®** - Knowledge Assessment Teaching Engine
- Best Practices for Programming:
  - Testing
  - Code Quality

CAMBRIDGE SPARK

Ice-breaker







CAMBRIDGE SPARK

We put the emphasis on **practice**

- Taught weekends:
  - develop intuition, ask questions, use “simple” datasets
- Practical homework exercises:
  - self-paced, remote support (slack channel), “realistic” data
- Cambridge Spark’s Goal:
  - All graduates equipped to join or form a Data Science team
- You will:
  - Learn fundamental techniques
  - Learn how to develop robust, version controlled code with a team
  - Learn how to communicate your results

CAMBRIDGE SPARK

The teaching team

- **15+ people**: experts with diverse backgrounds
- **Normal format**: short lectures followed by practicals
- **Questions**: ask away, but please be considerate in lectures -- flow is important
- **Nuanced questions**: Slack is your friend!
- **Feedback**: https://goo.gl/forms/qfzRUVwP1dXMa8Kg1

CAMBRIDGE SPARK

Module homework

Expect ~10 hours of self-study between teaching sessions

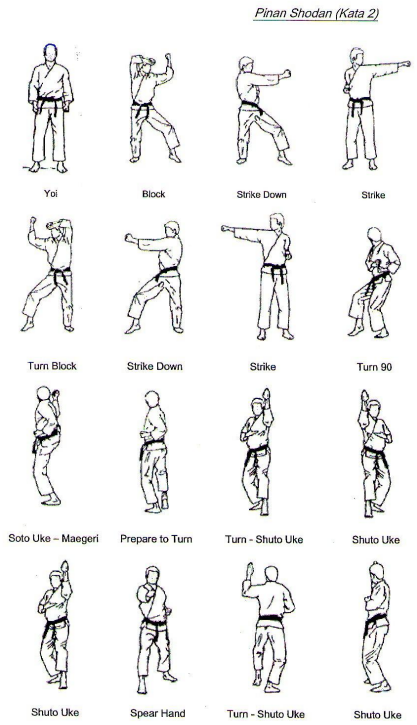
Two forms:

1. Defined coding exercises with **K.A.T.E.®**
2. More free open ended questions

CAMBRIDGE SPARK

Defined coding exercises: “kata”

1. Coding skill development
2. **K.A.T.E.®** gives feedback:
    a. Test cases: meets spec? edge cases?
    b. Code style
3. View your progress on the dashboard



CAMBRIDGE SPARK

Open ended questions

- Task briefing + dataset:
    - Hope you practiced your kata!
- Self-driven
- Presentations and discussions

CAMBRIDGE SPARK

Capstone project

- 6 weeks of work (50+ hours)
- Industry partners provide real problems…
- ...but you can do your own personal project instead
- You get a 1 to 1 supervisor to help
- Group presentation
- Effort in $\propto$ benefit gained

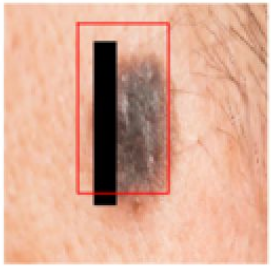
CAMBRIDGE SPARK

An example project

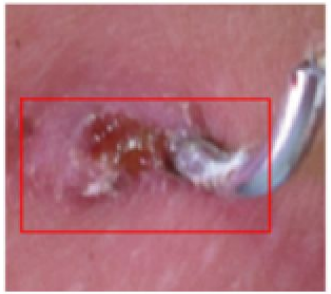
https://blog.cambridgespark.com/early-infection-detection-using-yolo-10d593970794

MELANOMA

MELANOMA WITH OCCLUSION



PIN SITE INFECTION



Images from Google chosen at random

CAMBRIDGE SPARK

Summary and key points

- Don’t study...**Do!**
- Read the docs
- Don’t stress!
- Use Slack

CAMBRIDGE SPARK

Programme overview

- First 4 sessions - A practical introduction to **Machine Learning**
  - Get to grips with python packages and K.A.T.E.
  - Data exploration and feature creation
  - Introduction to classification, regression, time series', and unsupervised learning
  - Model evaluation, regularisation, and selection
- Middle 2 sessions - Scaling up and out
  - Big data considerations
  - Cloud computing
  - Databases
- Final 4 sessions - Advanced topics
  - Ensemble methods e.g. Random Forests and Gradient Boosting
  - Neural Networks & Deep Learning
  - Natural Language Processing
  - Recommender Systems & Interpretability

CAMBRIDGE SPARK

What now?

- Software and packages:
  - Git - version control
  - Conda - package management
  - Jupyter - python IDE
  - Numpy
  - Pandas
- Mathematics:
  - Linear Algebra
  - Probability
  - Optimisation
- **K.A.T.E.®**
- Coding best practices

CAMBRIDGE SPARK