Presentation Slides:

For INSTRUCTOR Use Only

Machine Learning

Licensed for personal use only for Fernando K <fernando_kruse@dell.com> from Machine Learning at Dell Brazil (QE) @ 2019-03-12

Prerequisites & Expectations

- Development background
  - Need to be comfortable programming - there are many labs
  - Our labs are in Python and Scala
    (If you don't know Scala, we will do a Scala Primer at the start)

- Basic knowledge of Linux

- We don't expect Math or Statistics background

Real World Datasets

- We will work on real world datasets such as
  - Uber trips data
  - Walmart shopping data
  - Netflix recommendation
  - Health studies
  - Credit card default data
  - Prosper loan data
  - NYSE stock data

- No old-worn-out datasets like Iris / Titanic …

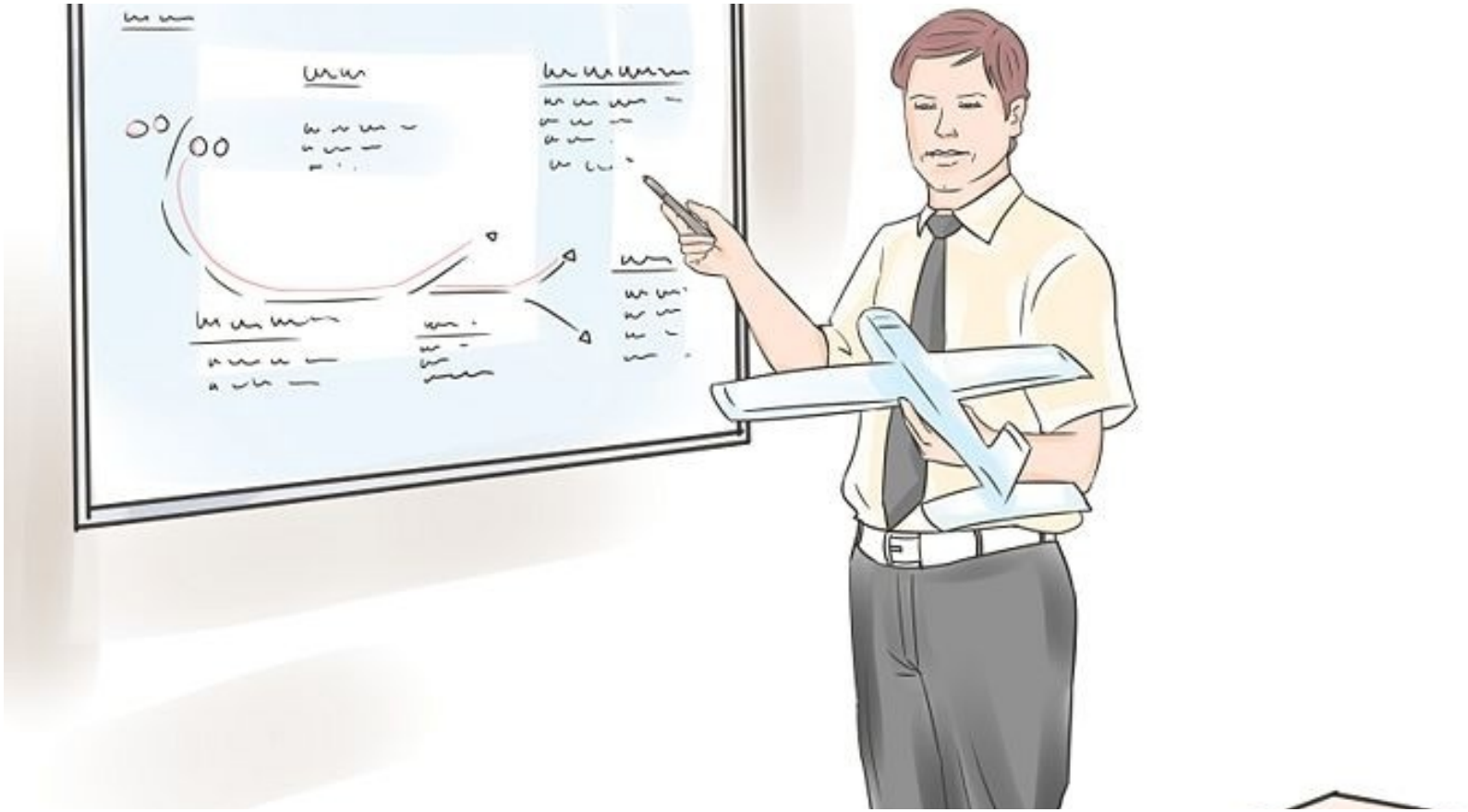
Workshop Overview

- **Day 1**
  - ML intro
  - ML concepts
  - Statistics primer
  - Data exploration & visualizing
  - Feature engineering
  - Linear regression
- **Day 2**
  - Logistic regression
  - SVM
  - Decision trees
  - Random forest
  - Naïve Bayes
- **Day 3**
  - Clustering
  - K-Means
  - PCA
  - Recommendations
- **Workshops**
  - Final day, solve a problem end to end
  - Daily workshops (time permitting)

Lots of Labs: Learn by Doing

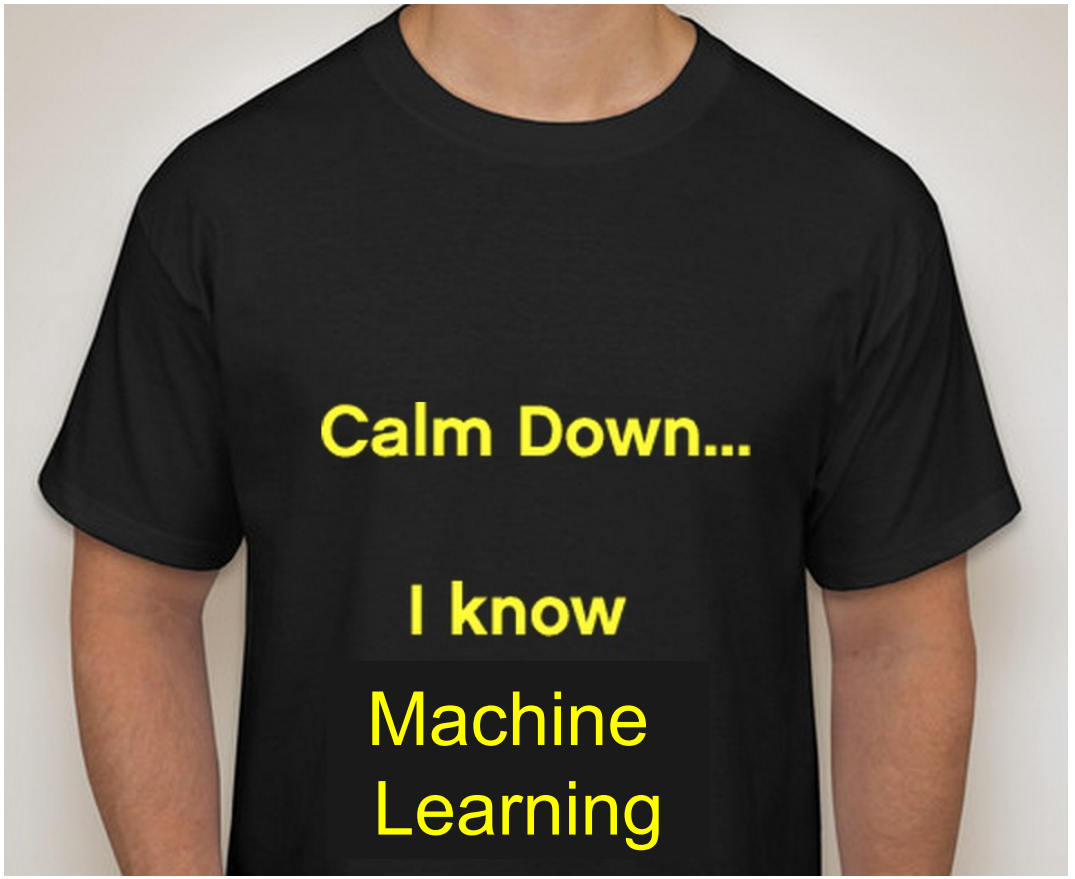
Analog: Learning to Fly…

Instruction



+ Flight Time

After The Class...



About You and Me

- About you
  - Your name
  - Your background (developer, admin, manager, etc.)
  - Technologies you are familiar with
  - Familiarity with Machine Learning
    (scale of 1 – 4 ; 1 – new, 4 – expert)
  - **Something non-technical about you!**
    (favorite ice cream flavor, hobby, etc.)

Recommended Resources

- Please see our handout

Class Logistics

- Instructor's contact
- Slides
  - For each session, slides will be emailed out or delivered via virtual classroom
- Labs
  - Lab files will be distributed
- Playground
  - Provided in the cloud

Licensed for personal use only for Fernando K <fernando_kruse@dell.com> from Machine Learning at Dell Brazil (QE) @ 2019-03-12

Typographic Conventions

- Code in the text uses a fixed-width code font, e.g.:

  `catalog: Catalog = new CatalogImpl`

  - Code fragments are the same, e.g. `catalog.speakTruth`
  - We **bold/color** text for emphasis
  - Filenames are in italics, e.g. *Catalog.scala*
  - Notes are indicated with a superscript number (1) or a **star ***
  - Longer code examples appear in a separate code box - e.g.

```
object TestApp {  // Basic Spark App (Scala)
  def main(args: Array[String]) {
        val sc = new SparkContext(
      new SparkConf().setMaster("local").setAppName("TestApp")
        val totalWords = sc.textFile("file")
                .flatMap(l => l.split(" ")).count()
        println ("# lines : " + totalWords)
    }
}
```

Questions?

- Any Questions?