Painless Predictions - Python

Classification:

Google: Given a search query, is a website **relevant** or **irrelevant**?

Facebook: Is this image a **face** or a **non-face**?

Is this user a good potential customer or no?

NON-Classification problems:

Machine translation, languages

The world is not binary (or really even in discrete classes at all)

Example: Gallup party registration poling:

Democrat = 31%

Republican = 26%

In the real world

Find some data with classes already attached.

Common algorithms (a way to solve a problem, not the problem itself):

* Logistic Regression
* Decision Tree methods (such as Random Forest)
* Support Vector Machines

Avoid over-fitting to the weirdness in the room.

Focusing on Regression

Creates a formula based on the independent (known) variables to predict the value of the dependent variable. Regression: Generating a curve from data instead of data from curves.

The logistic function is an equation that starts at 0 and ends at 1.

Dependent: two classes (yes and no)

Independent: multiple classes (age)

Scale the data: continuous data can be arbitrarily large or small, so hone it in to avoid crazy outliers.

Precisions can be false negatives, and Recall can have false positives