

[Outline](#) > 2. Data And Features > Lecture: Determining Features > Video

Video

[Bookmark this page](#)

Determining Features

With a challenge in mind you want to solve, you know that you need to collect a lot of samples along with features that describe them. This is the prerequisite for machine learning. You now also know these features can be continuous numeric values, or they can be categorical. But how exactly should you go about choosing features? And what's more important to focus on—adding additional features or collecting more samples?

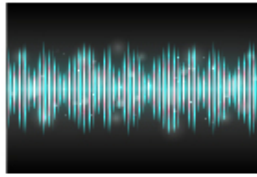
These are reasonable questions everyone has when they start amassing the data they need to solve an issue. The answer is, *it depends*. Just as in the example of Angie & Craig's lists mentioned in the Machine Learning section, your own intuition about the problem being tackled should really be the driving force behind what data you collect. The only unbreakable rule is that you need to ensure you collect as many features and samples as you possibly can.

If you ever become unsure which one of the two you should focus on more, concentrate on collecting more samples. At least during collection, try to make sure you have more samples than features because some machine learning algorithms won't work well if that isn't the case. This is also known as the curse of dimensionality. At its core, many algorithms are implemented as matrix operations, and without a greater than or equal number of samples to features, a fully-formed system of independent equations cannot be made. You can always create *more* features based off of your existing ones. But creating *pseudo-samples*, while not impossible, might be a bit more difficult.

INPUT

FEATURES

INTELLIGENCE


 $[X_0, X_1, X_2, \dots, X_i]$


"What in the world...?"

"It is a curious thing, Harry, but perhaps those who are best suited to power are those who have never sought it. Those who, like you, have leadership thrust upon them, and take up the mantle because they must, and find to their own surprise that they wear it well."


 $[X_0, X_1, X_2, \dots, X_i]$


—Albus Dumbledore


 $[X_0, X_1, X_2, \dots, X_i]$


Cookie Monster

© All Rights Reserved



© 2012-2017 edX Inc. All rights reserved except where noted. EdX, Open edX and the edX and Open EdX logos are registered trademarks or trademarks of edX Inc. | 粤ICP备17044299号-2

POWERED BY
OPENedX®

