

Where to Start When You're New to Machine Learning

Alyssa Batula

PyData Indy

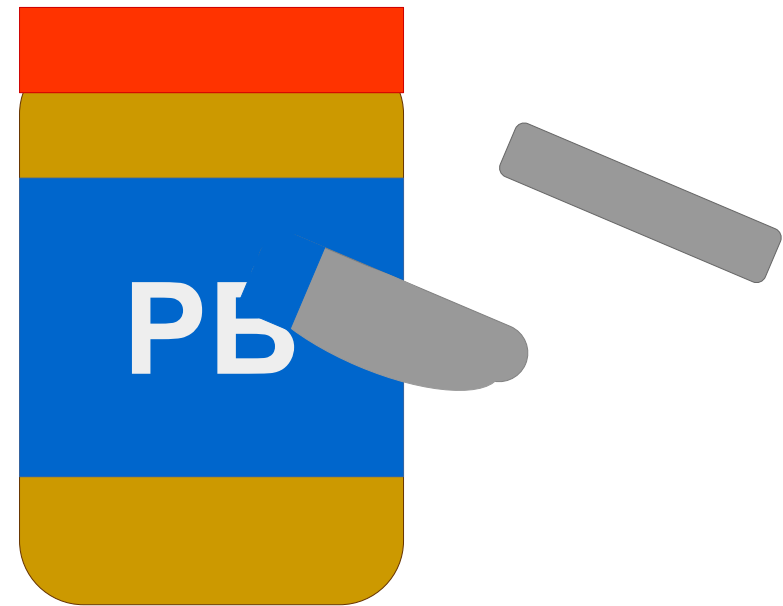
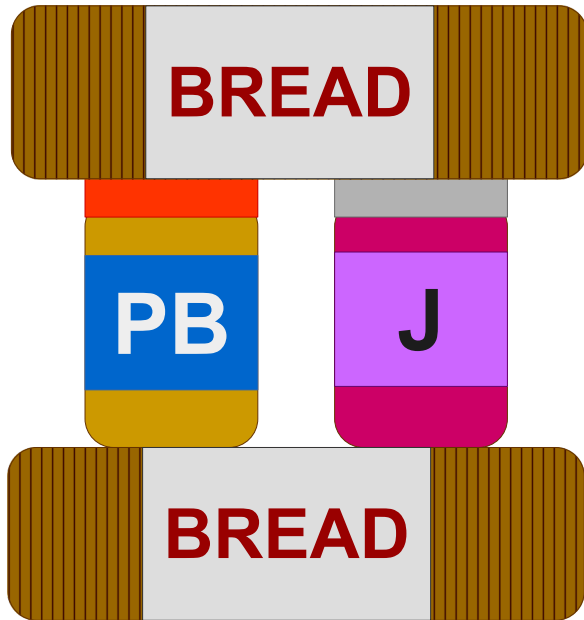
October 2019

What is Machine Learning?

“Machine learning is the science of getting computers to act without being explicitly programmed.”

~Andrew Ng

Teach a Computer to Make a Sandwich



First Steps

1. Change your mindset

Machine Learning Mindset

- Data management instead of code
- Unknowns and experiments instead of common solutions
- New definitions for determining good solution

First Steps

1. Change your mindset
2. Determine if machine learning is good fit

First Steps

1. Change your mindset
2. Determine if machine learning is good fit
3. Prepare your data

Preparing Dataset

- .Enough data

- .Accurate

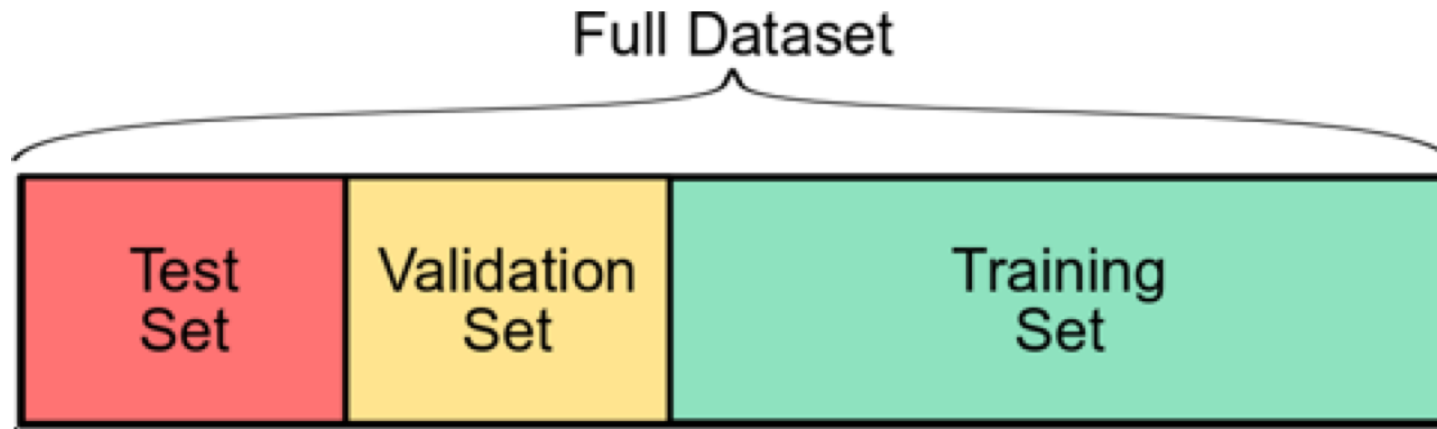
- .Relevant

- .Complete

What kind of data do you have?

- Labeled Categories → Classification
 - Cat/dog, true/false, red/green/blue, etc
- Number/value → Regression
 - Price, temperature, time
- No labels → Clustering/unsupervised learning
 - Data exploration

Training, Validation, and Testing



Training Set - Used to train a machine learning algorithm

Validation Set - (Optional) Used to compare algorithms

Testing Set - Used to evaluate final algorithm

Train/Test Split in Code

```
from sklearn.model_selection import train_test_split

# X - examples
X = [[3, 0], [9, 3], [4, 2], [2, 3], [8, 9]]

# y - labels
y = [0, 0, 0, 1, 1]

(train_X, test_X,
 train_y, test_y) = train_test_split(X, y,
test_size=0.25)
```

A Closer Look at the Data

```
# Get your data
X = [[3, 0], [9, 3], [4, 2],
      [2, 3], [8, 9]]
y = [0, 0, 0, 1, 1]
```

Machine Learning in Code

```
# Import from scikit-learn
from sklearn.model_selection import
train_test_split
from sklearn import svm

# Get your data
X = [[3, 0], [9, 3], [4, 2], [2, 3], [8, 9]]
y = [0, 0, 0, 1, 1]
train_X, test_X, train_y, test_y =
train_test_split(X, y, test_size=0.25)
```


Evaluate The Model

```
from sklearn.metrics import accuracy_score

pred_y = ml_model.predict(test_X)

accuracy = accuracy_score(test_y, pred_y)
```

Evaluation Metrics

.Classification

- Accuracy
- Precision
- Recall

.Regression

- R^2
- Mean squared error

First Steps Recap

- .Prepare for a new mindset
- .Decide if machine learning is the right solution
- .Clean and prepare your data
- Use your data to choose an algorithm
- .Try it out!

Resources

- [Podcast: Practical AI](#)
- [Scikit-learn Documentation](#)
- [3Blue1Brown Neural Network Video](#)
- [Neural Networks & Deep Learning](#)
- [PythonProgramming.net](#)
- [Coursera: Structuring Machine Learning](#)