**Regression:**

Read before class the following post “Selecting the best Machine Learning algorithm for your regression problem” :

[**https://towardsdatascience.com/selecting-the-best-machine-learning-algorithm-for-your-regression-problem-20c330bad4ef**](https://towardsdatascience.com/selecting-the-best-machine-learning-algorithm-for-your-regression-problem-20c330bad4ef)

Make notes on the following concepts

* No free lunch theorem
* Linear vs nonlinear
* SGD Stochastic Gradient Descent
* Neural Networks
* Decision Trees

**Regressions in scikit:**

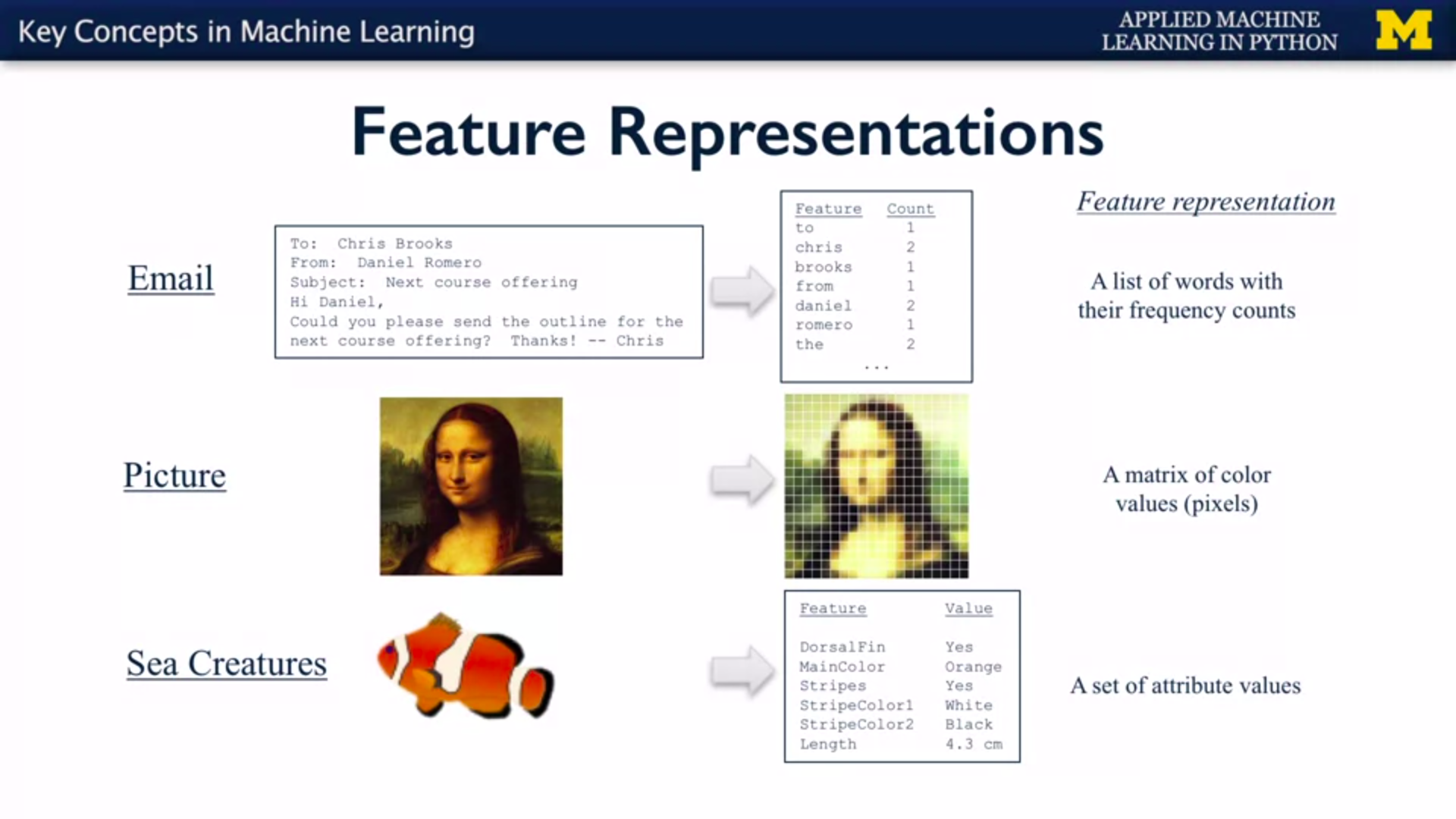
Algorithms that have the word regression of on their name (and actually do classification)

[**https://scikit-learn.org/stable/supervised\_learning.html**](https://scikit-learn.org/stable/supervised_learning.html)

**Linear Regression:**

1. Representing data in functions.

Slide taken from: <https://www.coursera.org/learn/python-machine-learning> by “Kevyn Collins-Thompson”

****

Use **Linear\_reg\_examples.docx for examples**

Make a graph using a small data set and watch the relationship between the dots and the function. (CPUs and Cost):

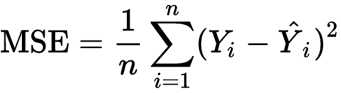
Plot data

Draw a line

1. Establishing **hypothesis** functions.

Activity**:** For a simple line try to find functions that fit the data by hand, then try to do so for a more complex function that does not have an exact hypothesis. Linaer\_reg\_examples.docx

Draw a line that you think would pass through most of the dots in the graph, Now write the function for that line.

1. Create **cost functions** (**loss functions**). Use MSE Mean Square Error

Compare your hypothesis with the real function and graph the error i.e. (the difference between the actual data and the predictions of your function).

1. **Optimization/fitting step:** If we have a cost function and a model, we can treat this as an optimization problem.

Math way. (Optimal) Matrix operations **b=(X'X)-1 X’y** but can be very expensive depending on the data set and has several restrictions (no sparsity, shape of the matrix, among others ...).

The Iterative way. (M.L.) (approximation through trial and error approach).

Finding good hypothesis functions (or treating ML problems as function optimization problems.)

Maximize and minimize functions.

Make a linear regression model for the following dataset, try to predict USD Based ISE: <https://archive.ics.uci.edu/ml/datasets/ISTANBUL+STOCK+EXCHANGE#>

**Exploratory Data Analysis**

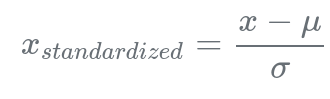
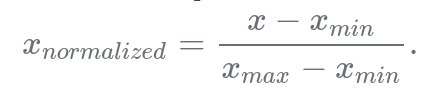
**Preprocessing (Transforming data)**

**Feature Scaling:**

<https://www.geeksforgeeks.org/ml-feature-scaling-part-2/>

In **scikit**:

<http://scikit-learn.org/stable/modules/preprocessing.html>

****

Follow example in transforming data **transforming data.ipynb**

**1 hot encoding (to use text labels in regressions)**

<https://medium.com/@michaeldelsole/what-is-one-hot-encoding-and-how-to-do-it-f0ae272f1179>

**word vectorization tougher, embeddings**

<https://www.analyticsvidhya.com/blog/2017/06/word-embeddings-count-word2veec/>