Intro to Machine Learning and Time Series Analysis

FinTech

Lesson 10.1

Class Objectives

Manipulate time series data sets in Pandas.

Identify the components of a time series.

Use moving average and exponentially-weighted moving average.

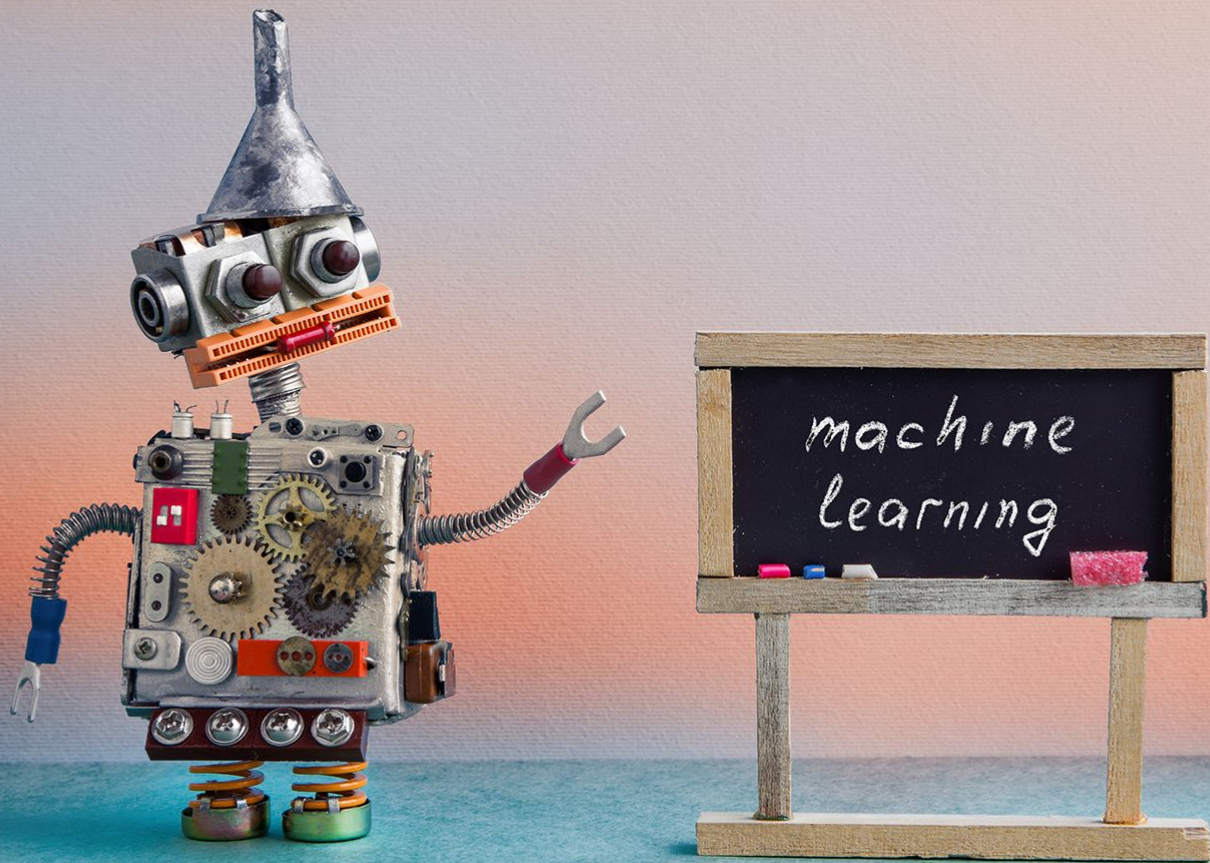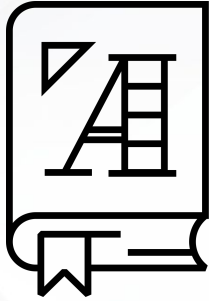Utilize Hodrick-Prescott filter to identify long-term trends.

Perform autocorrelation on time series data.

Mysticism of Machine Learning

So It Begins...

Machine learning is a programming approach allowing applications to learn from their inputs and make adjustments based on their outputs.

In a nutshell, **machine learning,** is developing statistical models that can make predictions or decisions on new data automatically.

Mysticism of Machine Learning

Machine Learning is used to predict:

Loan eligibility, foreclosure rates, and fraud.

Disease diagnosis and prognosis.

Consumer segmentation and clustering.

Presidential election results.

Natural disaster and planetary climate impacts.

Mysticism of Machine Learning

Machine Learning Models

01 **Libraries**

Models for machine learning are provided in libraries, just like other code we've used.

02 **Pipelines**

All machine learning pipelines use a **Model -> Fit-> Predict** paradigm. Once the model is fit, it can be used to make predictions.
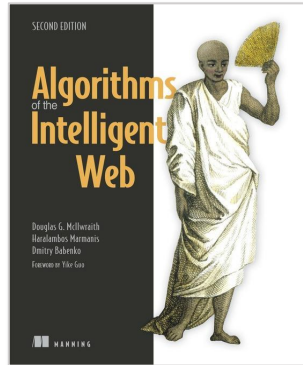
Mysticism of Machine Learning

> **Intelligent algorithms** are ones that use data to modify its behavior. Intelligent algorithms differ in that they can change their behavior as they run, often resulting in a user experience that many would say is intelligent.

—*Algorithms of the Intelligent Web, Second Edition*

**Algorithms of the Intelligent Web, Second Edition**
by *Douglas G. McIlwraith Haralambos Marmanis Dmitry Babenko*
Publisher: Manning Publications
Release Date: August 2016

Mysticism of Machine Learning

Intelligent algorithms are ones that respond to data such that the algorithm gets better. It effectively "evolves." The decision is no longer deterministic given the event.

Data

Event

Intelligent Algorithm

Decision
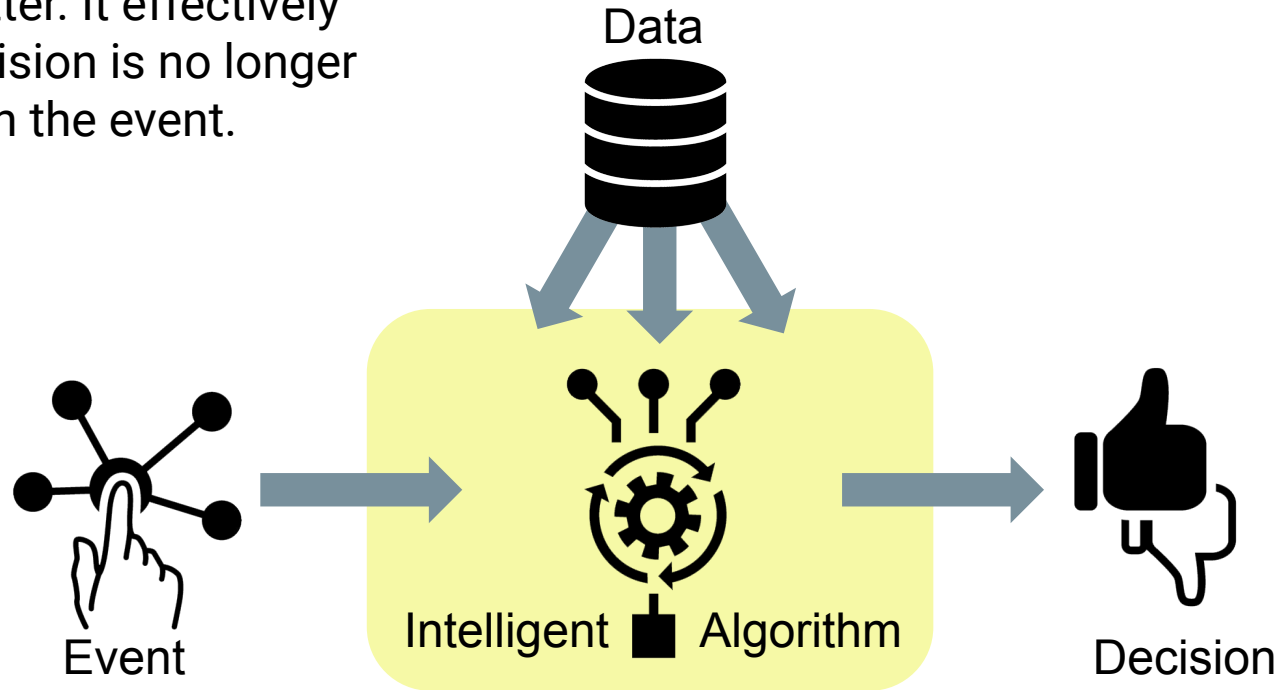
Mysticism of Machine Learning

Machine Learning
Capability of software to generalize phenomena (past or future) based on past experience

Predictive Analytics
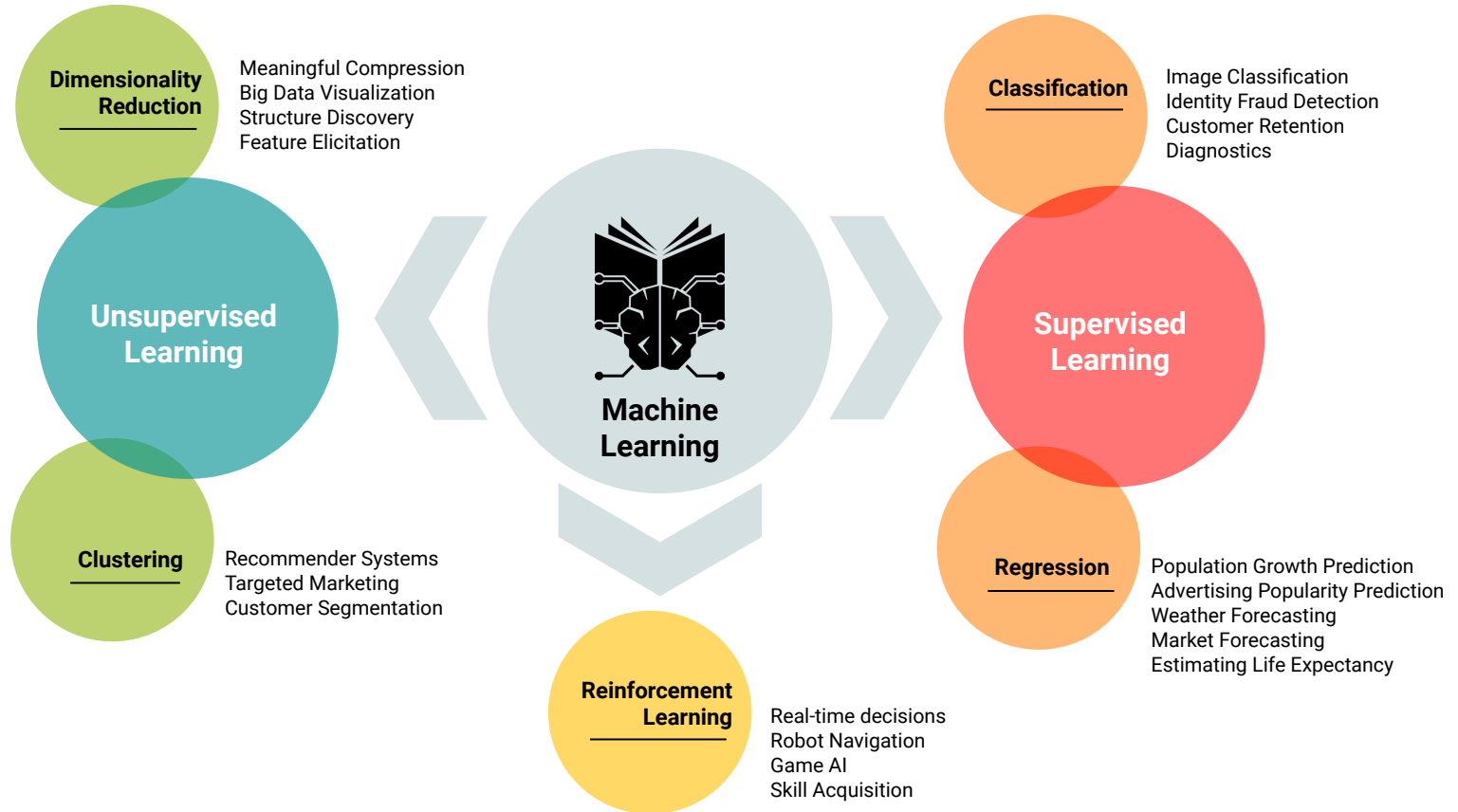Capability of software to predict future outcomes based on historical data

Artificial Intelligence
Software (and machines) that have a series of options to achieve a particular goal
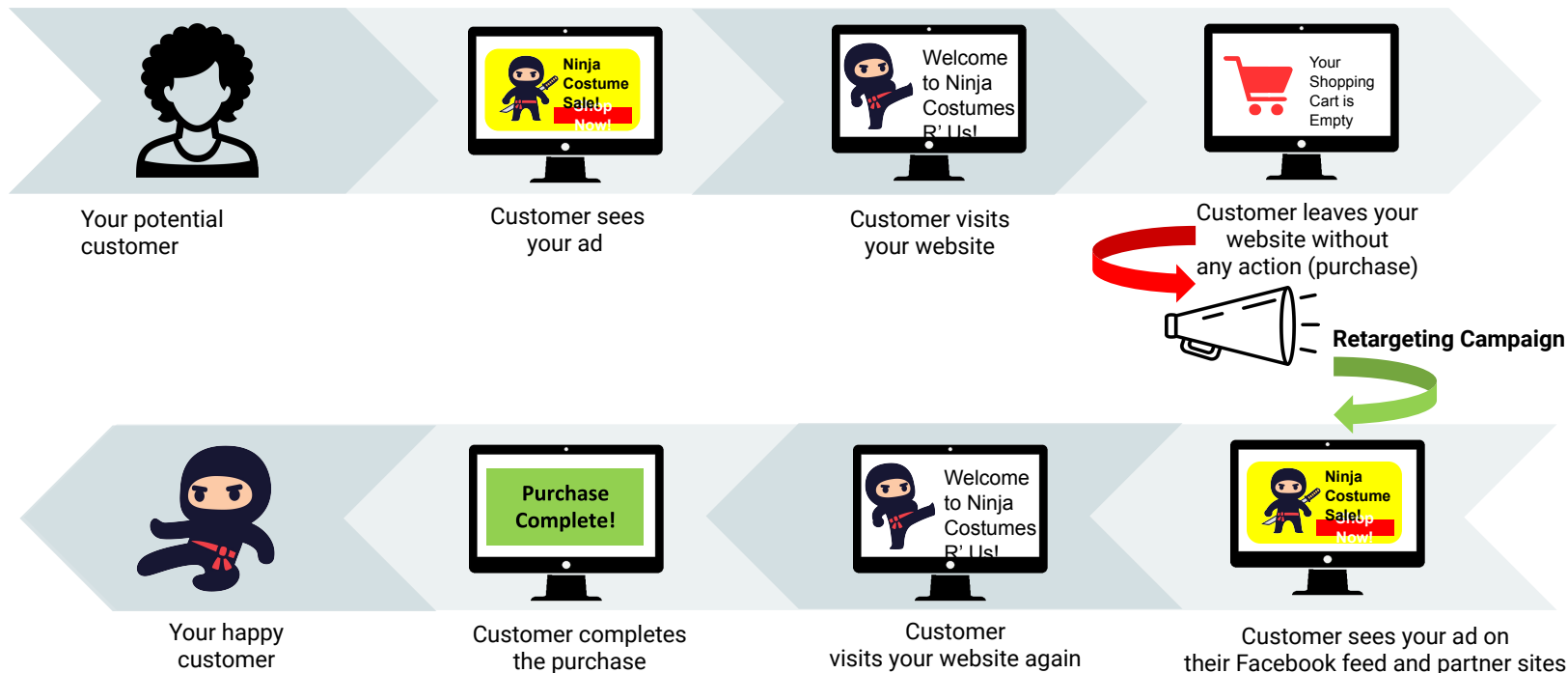
Mysticism of Machine Learning

Dimensionality Reduction
Meaningful Compression
Big Data Visualization
Structure Discovery
Feature Elicitation

Unsupervised Learning

Clustering
Recommender Systems
Targeted Marketing
Customer Segmentation

Machine Learning

Reinforcement Learning
Real-time decisions
Robot Navigation
Game AI
Skill Acquisition

Classification
Image Classification
Identity Fraud Detection
Customer Retention
Diagnostics

Supervised Learning

Regression
Population Growth Prediction
Advertising Popularity Prediction
Weather Forecasting
Market Forecasting
Estimating Life Expectancy

Mysticism of Machine Learning

Predictive Analytics

Your potential customer

Customer sees your ad

Customer visits your website

Customer leaves your website without any action (purchase)

Retargeting Campaign

Your happy customer

Customer completes the purchase

Customer visits your website again

Customer sees your ad on their Facebook feed and partner sites

Mysticism of Machine Learning

Artificial Intelligence

Instructor Demonstration
Review Homework

Time Series Basics

Time Series Basics

Working with time series data requires a return to the basics

Data needs to be sliced and diced at various time frequencies in order to analyze data points as a time series

E.g. day, week, month, year

6

Pandas DateTimeIndex index can be used to help with this

`df.loc[2019]`

Time Series Basics

The Pandas resample function can also be used to slice and dice data, once a DateTimeIndex has been created

```
weekly = df['Close'].resample('W').mean()
```

Instructor Demonstration

Time Series Basics

Activity: Time Series Basics

In this activity, you will practice the basics of time series manipulation in pandas.
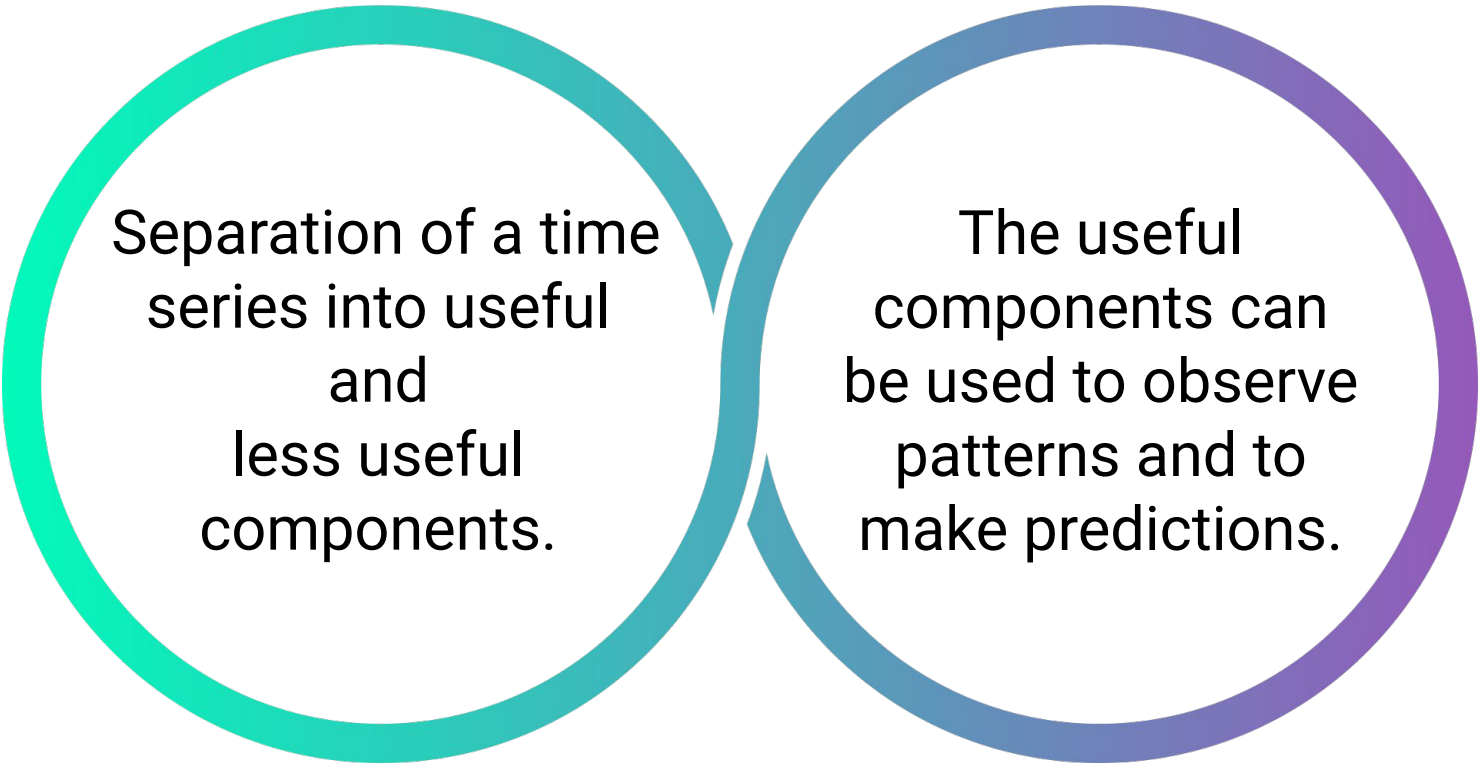
Instructions sent via Slack.

Suggested Time:
15 minutes

Time's Up! Let's Review.

Time Series Decomposition

Time Series Decomposition

Separation of a time series into useful and less useful components.

The useful components can be used to observe patterns and to make predictions.

Time Series Decomposition

Components

01 **Level:** What is the average value of the series?

02 **Trend:** Is there an overall direction of movement?

03 **Periodicity:** Do patterns occur in cycles?

04 **Residual:** How much noise exists in the data?

Instructor Demonstration
Time Series Decomposition

Exponentially-Weighted Moving Average

Exponentially Weighted Moving Average (EWMA)

EWMA is an approach used to "denoise" or "smooth" out time series data so that trends and predictions can be made

01

EWMA involves calculating the average of the last n prices

02

Weights are added to the averages based on the recency of the data

- Recent data is weighted more heavily
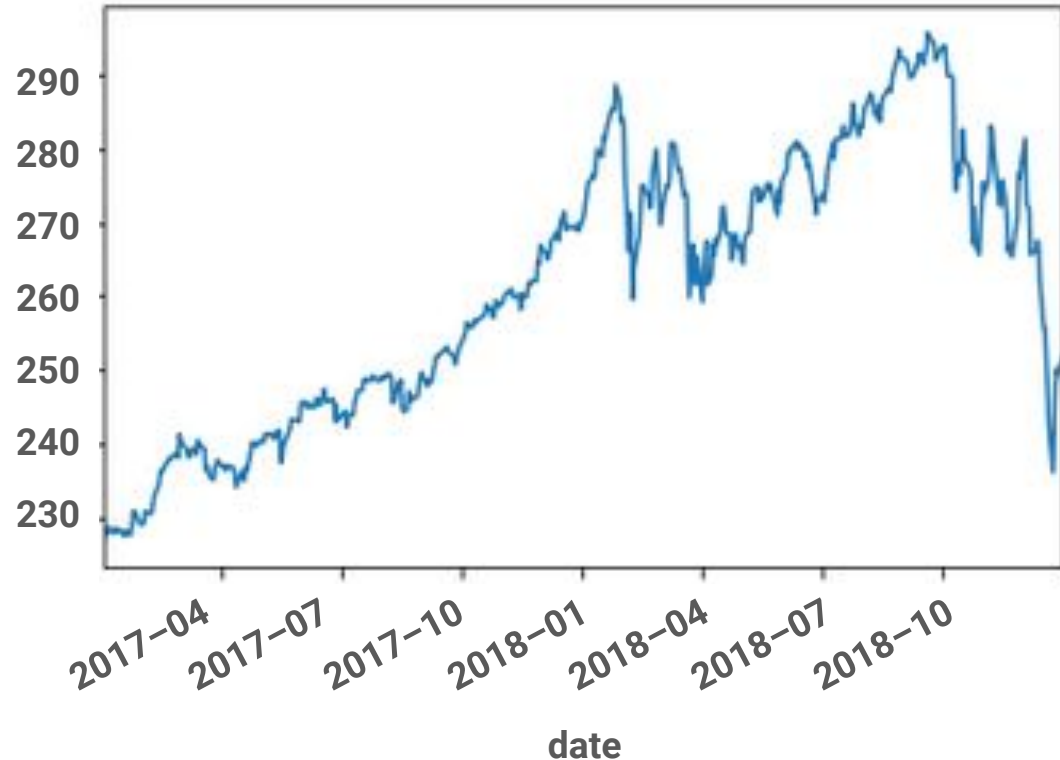- Weighting decreases exponentially for previous prices/time periods

03

Requires past average values to be stored in memory

Exponentially Weighted Moving Average (EWMA)

EWMA is used to highlight trends and illustrate the price trajectory for an investment

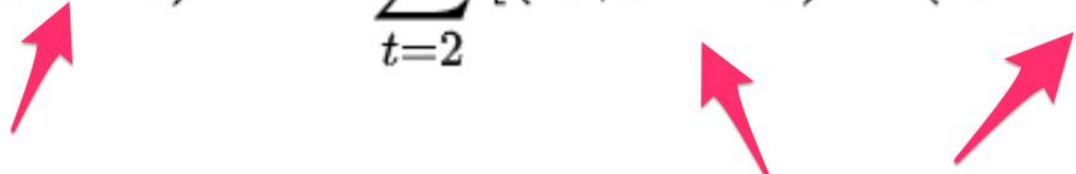In which direction is the price moving?

Hodrick-Prescott Filter

Hodrick-Prescott Filter: A mathematical function that separates a time series into trend and non-trend components.

Hodrick-Prescott Filter

- Filters out short-term fluctuations.

- Decomposes a time series into trend and non-trend components.

Hodrick-Prescott Filter

$$\min_{\tau} \left(\sum_{t=1}^{T} (y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} \left[(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1}) \right]^2 \right)$$

time series value - trend = cyclic element

Difference in trend over time = volatility

Instructor Demonstration
EWMA and Hodrick-Prescott Filter

Activity: You've got a FRED

In this activity, you will use the Hodrick-Prescott filter to identify macroeconomic trends in the United States in the period from 2004 to 2010.
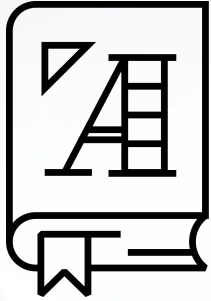
Instructions sent via Slack.

Suggested Time:
15 minutes

Time's Up! Let's Review.
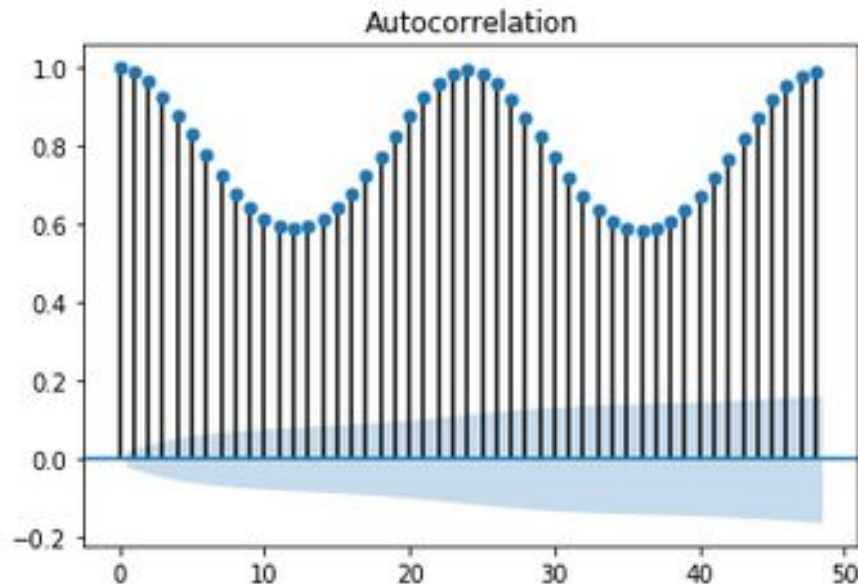
Autocorrelation

Autocorrelation is a measure of how closely current values correlate with past values

For example, **autocorrelation** is used to determine to what extent today's prices correlate with yesterday's prices

Autocorrelation

Hourly temperature is a great example of the concept of autocorrelation.

- ● What's the temperature at noon today?

- ● It is likely we'll get good information by looking at what the temperature was at noon yesterday (24 hours, or "lags," ago).

Autocorrelation

Instructor Demonstration
Autocorrelation

Activity: Euro ETFs

In this activity, you will examine a time series of bid-ask spreads of an ETF for autocorrelation.

Instructions sent via Slack.

Suggested Time:
15 minutes

Time's Up! Let's Review.

Questions?