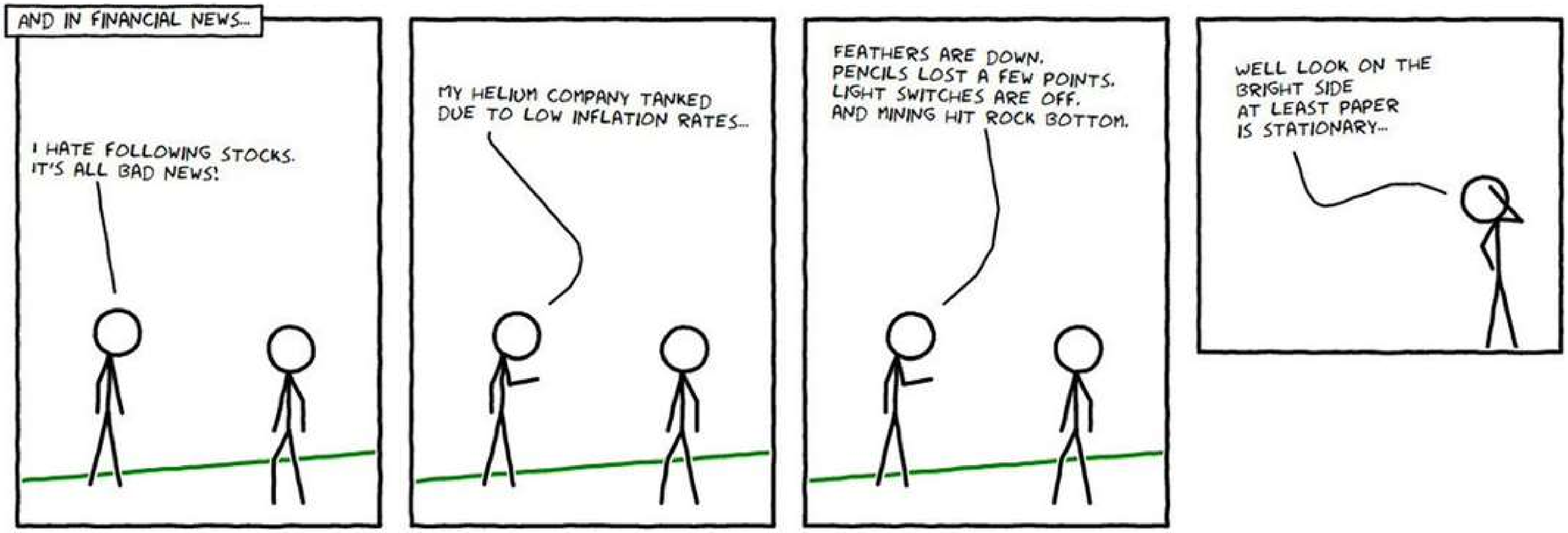
COMPUTER BASED PATTERN RECOGNITION IN QUANTITATIVE TRADING

Class of 2024


AND IN FINANCIAL NEWS...
I HATE FOLLOWING STOCKS.
IT'S ALL BAD NEWS!
MY HELIUM COMPANY TANKED
DUE TO LOW INFLATION RATES...
FEATHERS ARE DOWN.
PENCILS LOST A FEW POINTS.
LIGHT SWITCHES ARE OFF.
AND MINING HIT ROCK BOTTOM.
WELL LOOK ON THE
BRIGHT SIDE
AT LEAST PAPER
IS STATIONARY...

RECALL

In our 1st class, we had discussed:

1. The different categories of quantitative trading strategies

2. Discussed the advantages of market neutrality and how to convert any equities portfolio to market neutral

3. Discussed statistical arbitrage, starting with Pairs Trading
   - We had utilized econometrics tools including distance measures, OLS and cointegration

IN THIS DECK:

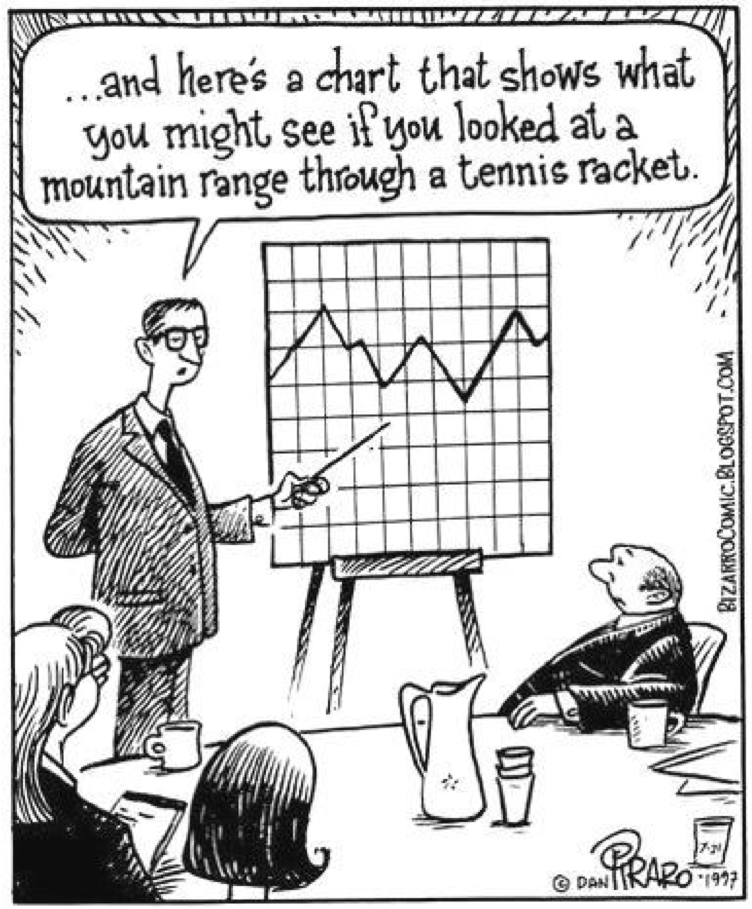
1. We will extend our toolbox to **evaluating time series heuristics** from technical analysis:
   - Overview of some short term market indicators
   - Use of kernel estimators and **support vector machines (SVMs)** for computer based automated recognition of visual patterns

2. Discuss implementation of pattern recognition via support vector machines in provided Python code [**basic SVM implementation over stock prices in Python**]

3. Discuss final project to **be presented by each group towards end of term**

TECHNICAL ANALYSIS

Astronomy or Astrology?


...and here's a chart that shows what you might see if you looked at a mountain range through a tennis racket.
BIZARROCOMIC.BLOGSPOT.COM
© DAN PIRARO · 1997

TECHNICAL ANALYSIS IS CONTROVERSIAL

"One of the greatest gulfs between academic finance and industry practice is the separation that exists between technical analysts and their academic critics. … It has been argued that the difference between fundamental analysis and technical analysis is not unlike the difference between astronomy and astrology. Among some circles, technical analysis is known as *voodoo finance*"

1st sentence from Lo, Mamaysky & Wang, "Foundations of Technical Analysis: Computation Algorithms, Statistical Inference and Empirical Implementation", The Journal of Finance Volume 55 Issue 4

TECHNICAL ANALYSIS IS CONTROVERSIAL

The attitude of academics towards technical analysis, until recently, is well described by Malkiel (1981):

> Obviously, I am biased against the chartist. This is not only a personal predilection, but a professional one as well. Technical analysis is anathema to the academic world. We love to pick on it. Our bullying tactics are prompted by two considerations: (1) the method is patently false; and (2) it's easy to pick on. And while it may seem a bit unfair to pick on such a sorry target, just remember: it is your money we are trying to save.

IS IT REALLY THAT BAD?

Technical analysis consists of a set of indicators based on historical time series data which generate buy/sell signals

The indicators are more often than not, self-contained. i.e. only time series for a stock S is used to generate signals for S. In many cases, no other information is used to generate signals for S, except for S's own time series

We can organize technical analysis indicators into 2 further categories:

1. Closed form formulas (e.g. RSI, MACD, etc)
2. Graphical / visual / chart patterns

Formula based Technical Indicators

FORMULA BASED

Easier to evaluate is a first class of indicators based on closed end formulae

As with majority of trading strategies, these can be categorized broadly into mean reversion and momentum indicators (or both, for some indicators)

SURVEY OF COMMON FORMULA BASED INDICATORS: STOCHASTIC OSCILLATOR

Stochastic Oscillator

Formula for this indicator is (Recent Close – Low(N)) / (High(N) – Low(N))

Where Low(N) and High(N) refers to the low and high for the last N periods

Intuitively, indicator shows where recent close lies in proportion to its recent range. Depending on N, this can be either a momentum or mean reversion indicator

In short run for small N (say N<=30), this indicator tends to be more suited for mean reversion. i.e. when assets are currently at the higher end of their recent range, they tend to fall (and vice versa).

In long run (for large N, say over a year), some analysts have used this indicator as a leading momentum signal

STOCHASTIC OSCILLATOR

Stochastic oscillator is an example of an oscillator. This is because its output value tends to "oscillate" within a well defined interval [0,1]

Because there are well defined boundaries, setting buy and sell signals are straightforward. In a mean reversion context, we might buy if the oscillator value is (say) <0.2 and sell if it is >0.8

We can interpret an oscillator value of (say) 0.2 as "the stock is in the bottom 20% of its recent trading range". Vice versa for a reading of 0.8

RSI

Another indicator in the oscillator category include:

RSI (relative strength index):

If prices are upward trending:

U = close(today) – close(previous), D = 0

Else:

U = 0, D = close(previous) – close(now)

Average U and D are then calculated using an n period moving average (either simple or exponential). The moving averages are used to compute:

RS = MA(U,n) / MA(D,n), and

RSI = 100 – 100/(1+RS)

RSI therefore takes values between 0 and 100. High RSI (e.g. >70) may indicate that price has moved up too quickly, and is therefore overbought. Similarly, a low RSI (<30) might indicate that the price is oversold

MACD

The MACD (moving average convergence / divergence) is calculated as the difference between 2 time series which are a function of asset prices. Specifically:

MACD = N_period EMA (price) – M_period EMA(price) where N << M

In practice, a common value for N is 12, while a common value for M is 26 and EMA stands for exponential moving average.

Exponential moving averages over a time series P(t) are computed recursively as:

EMA(t) = sigma*P(t) + (1- sigma)*EMA(t-1)

MACD values therefore fluctuate around 0. The intuition is that it should measure momentum in the market; specifically, a high positive reading of MACD may indicate that upward momentum is significant

ARMS INDEX (TRIN)

Rare among technical indicators, the ARMS index uses information from multiple assets rather than just a single stock. The formula is:

TRIN = (Number of advancing stocks (t) / number of declining stocks (t)) /(total volume of all advancing stocks (t) / total volume of all declining stocks (t))

Background: The index is intended as a measure of market sentiment by measuring the gap between market supply and market demand. It is intended to be used on a contrarian basis by determining when the overall market might change direction.

The index fluctuates around 1. Overall, if numerator > denominator (TRIN >1), this is a bearish signal. i.e. the recent market advance is not supported by strong volume. Vice versa for TRIN < 1

AVERAGE TRUE RANGE (ATR)

Rather than being a mean reversion (contrarian) or momentum (leading) indicator, the average true range (ATR) attempts to **determine entire range of an asset price for that period**

Define:

- **True Range (TR) over a day** as Max((High – Low), Abs(High – Close price), Abs (Low – Close price)) for a period

- Average True Range (ATR) = simple average of all **TR over a time period n**

Table III

Standard Test Results for the Fixed-Length Moving (FMA) Rules

Results for daily data from 1897–1986. Cumulative returns are reported for fixed 10-day periods after signals. Rules are identified as (short, long, band) where short and long are the short and long moving averages respectively, and band is the percentage difference that is needed to generate a signal. "N(Buy)" and "N(Sell)" are the number of buy and sell signals reported during the sample. Numbers in parentheses are standard t-ratios testing the difference of the mean buy and mean sell from the unconditional 1-day mean, and buy-sell from zero. "Buy > 0" and "Sell > 0" are the fraction of buy and sell returns greater than zero. The last row reports averages across all 10 rules.

Test	N(Buy)	N(Sell)	Buy	Sell	Buy > 0	Sell > 0	Buy-Sell
(1, 50, 0)	340	344	0.0029	−0.0044	0.5882	0.4622	0.0072
			(0.5796)	(−3.0021)			(2.6955)
(1, 50, 0.01)	313	316	0.0052	−0.0046	0.6230	0.4589	0.0098
			(1.6809)	(−3.0096)			(3.5168)
(1, 150, 0)	157	188	0.0066	−0.0013	0.5987	0.5691	0.0079
			(1.7090)	(−1.1127)			(2.0789)
(1, 150, 0.01)	170	161	0.0071	−0.0039	0.6529	0.5528	0.0110
			(1.9321)	(−1.9759)			(2.8534)
(5, 150, 0)	133	140	0.0074	−0.0006	0.6241	0.5786	0.0080

PRIOR RESEARCH ON EFFECTIVENESS OF FORMULAE BASED TRADING

Brock, Lakonishok, LeBaron (Journal of Finance 'Dec 92) find statistically significant results for trading based on moving average rules

OVERALL

We note that there are more than 100 indicators existing

A potential trader is encouraged to evaluate each indicator from the perspective of:

1. Economic reasoning / intuition behind each indicator. What is their rationale?
2. Potential overlaps / duplication between indicators with minor differences in mathematical expressions
3. Adopting a data driven approach to determine effectiveness. This can be via:
   - Surveying results from other researchers (e.g. in academic journals), if possible
   - Verifying with own backtest

[PRACTICUM ON FORMULAE INDICATORS]

- We will transition at this point to overview provided Python code to:
    1. Implement selected indicators in Python
    2. Backtest to obtain systematic cumulative PnL

- Please refer to uploaded Python code

Pattern based Indicators

IS IT REALLY THAT BAD?

One criticism of technical analysis, especially of "graphical / visual / chart pattern" category of indicators is focused on idea that it is not replicable or consistent. i.e. it may be *subjective*

For instance, two traders looking at same chart, and employing identical technical analysis techniques may arrive at different conclusions. This is partly a result of visual nature of the exercise

EXAMPLE OF SUBJECTIVITY

Two traders looking at this chart employing technical analysis may arrive at different conclusions.

One might see a double bottom

The other may not

This inconsistency / subjectivity makes it more difficult to test hypotheses based on technical analyses rigorously

SYSTEMATIZING TECHNICAL ANALYSIS IS NECESSARY BEFORE WE CAN EVALUATE IT

Lo, Mamaysky & Wang ('00) propose using a computer based pattern recognition algorithm to systematize pattern identification

They list a class of statistical estimators called smoothing estimators, specifically kernel regression, which they use to automate chart based technical analysis

KERNEL REGRESSIONS

Kernel regressions can be used to find non-linear relations between two random variables, X and Y

It allows us to describe non linear variations in a time series as a mathematical equation

Once we have a kernel regression representation of a time series, we proceed to look for maxima and minima along the kernel regression representation of this time series. This is because most chart patterns can be expressed as a sequence of maxima and minima

KERNEL REGRESSIONS AND SMOOTHING

For our use, the kernel regression attempts to fit a mathematical expression to describe a non linear curve for stock prices in the rolling window

Yellow line is our 'fitted' and 'smoothed' curve, $P(t) = m(t) + \varepsilon_t$

KERNEL REGRESSIONS AND SMOOTHING

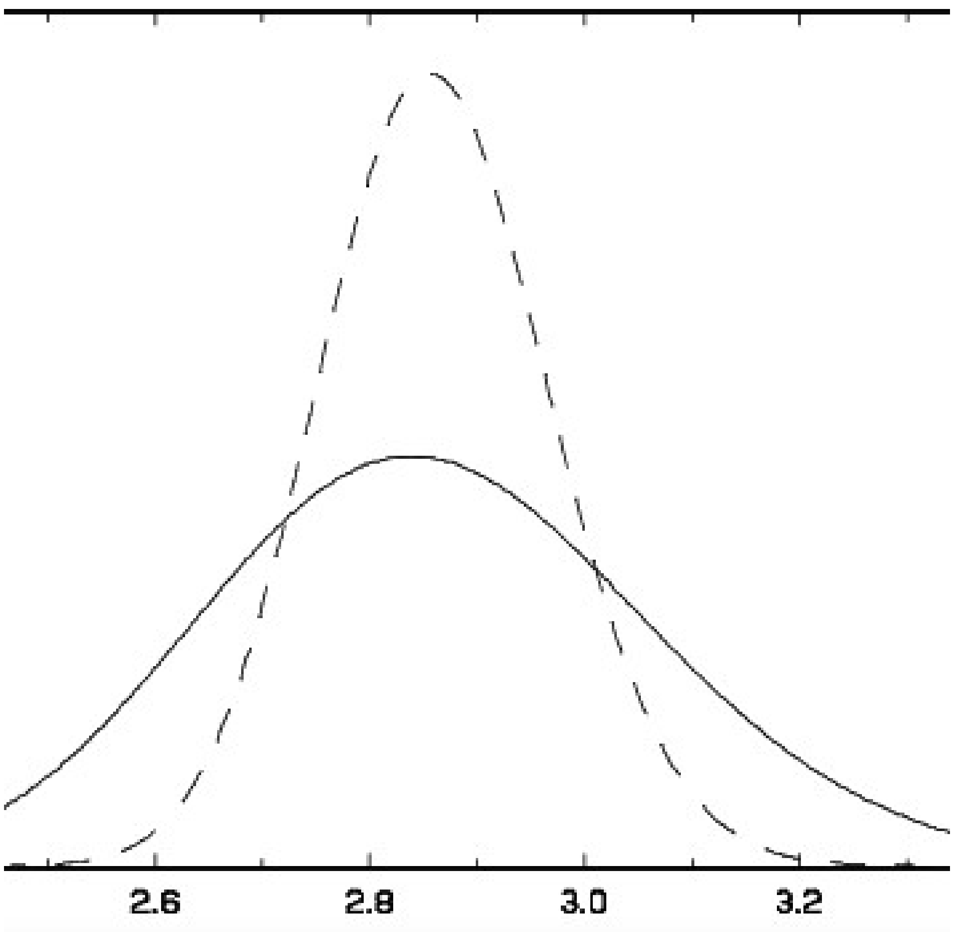
First, for each data point we have, (P,t), we compute a probability density function around that datapoint.

We use a gaussian kernel in this case: $$K(x) = \frac{1}{h\sqrt{2\pi}} e^{-0.5\left(\frac{x-xi}{h}\right)^2}$$

Width of probability density function is determined by h, which we term the bandwidth, and extent of smoothing

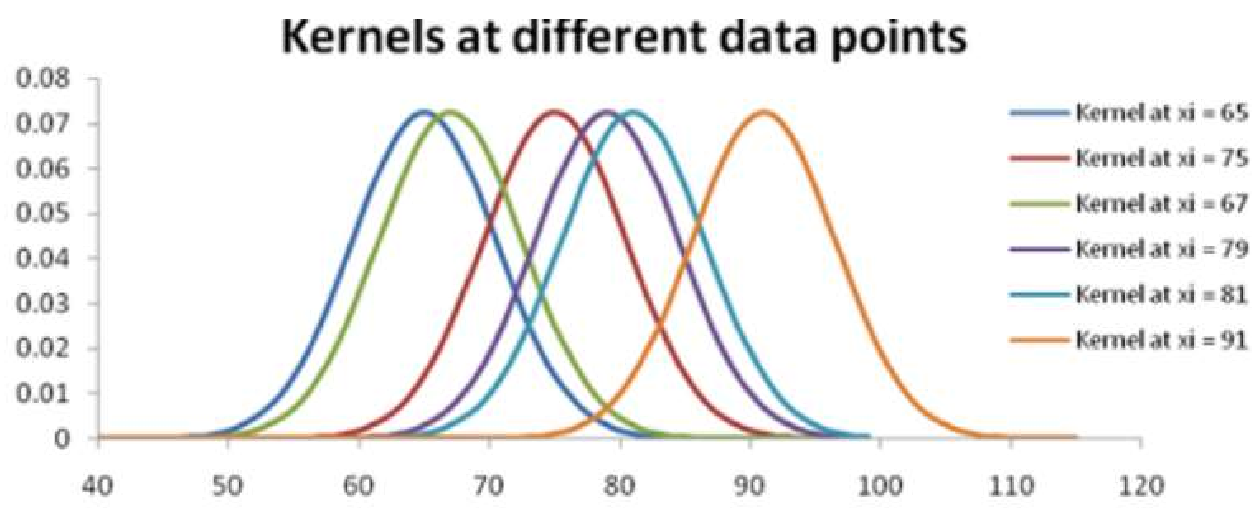
KERNEL REGRESSIONS AND SMOOTHING

A larger h leads to a ‘wider’ curve, which will in turn result in greater smoothing in our kernel regression



KERNEL REGRESSIONS

For each datapoint, we fit a probability density function



Source: https://towardsdatascience.com/kernel-regression-made-easy-to-understand-86caf2d2b844

KERNEL REGRESSIONS

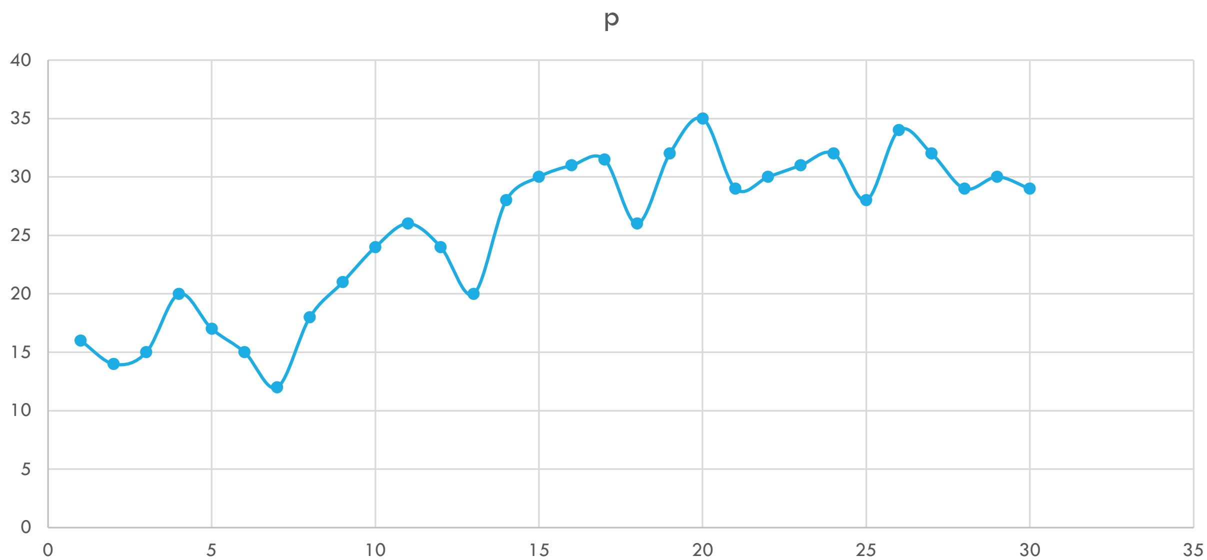
The predicted curve at each point **on t axis is weighted average of all datapoints at time t**

The weight of each datapoint at time t will be the **value of its probability density function at time t,** normalized by the sum of all the pdfs at t

We will go over a **worked example** in the next few slides

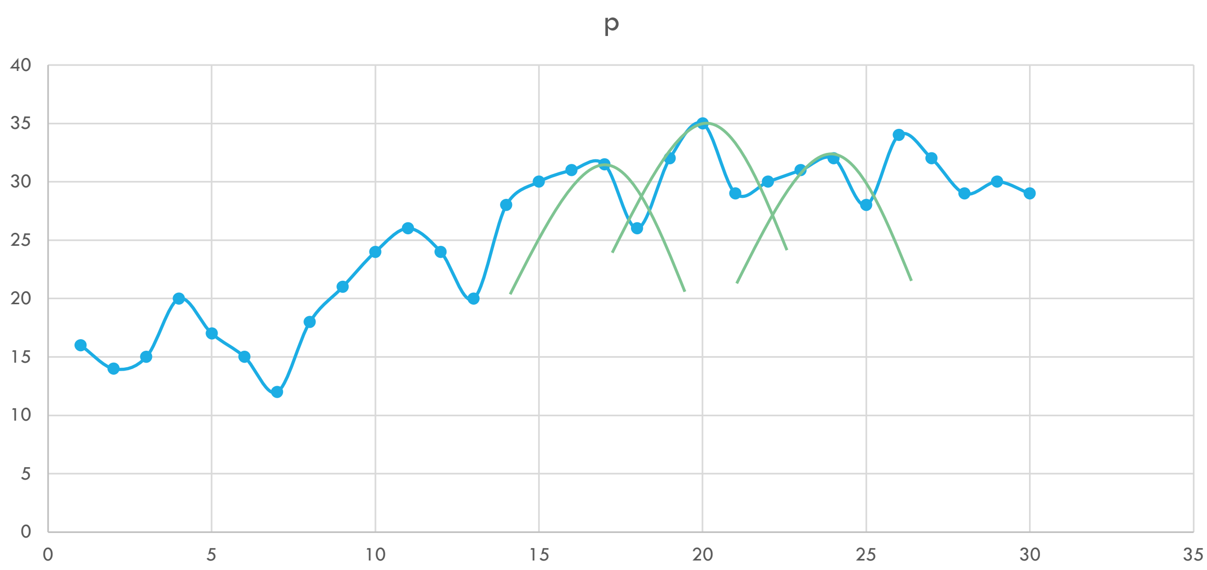
KERNEL REGRESSIONS

Assume we have the following price series **with 30 datapoints in our rolling window:**



KERNEL REGRESSIONS

We fit gaussian kernels to each datapoint (p,t). This ultimately gives us **30 kernels in each rolling window**



For ease of exposition on our next slide, label each of our 30 datapoints as p1, p2, p3, …p30

KERNEL REGRESSIONS

t	Pdf value of p1 at this value of t	Pdf value of p2 at this value of t	Pdf value of p3 at this value of t	…	Pdf value of p30 at this value of t
1	0.08	0.06	0.05		0.0005
2	0.07	0.07	0.08		0.001
3	0.05	0.06	0.09		0.003
4	0.03	0.03	0.08		0.005
…					
…					

KERNEL REGRESSIONS

Normalized kernel values

t	Pdf value of p1 at this value of t	Pdf value of p2 at this value of t	Pdf value of p3 at this value of t	…	Pdf value of p30 at this value of t	Sum across t
1	0.08/0.3	0.06/0.3	0.05/0.3		0.0005/0.3	0.3
2	0.07/0.4	0.07/0.4	0.08/0.4		0.001/0.4	0.4
3	0.05/0.2	0.06/0.2	0.09/0.2		0.003/0.2	0.2
4	0.03/0.5	0.03/0.5	0.08/0.5		0.005/0.5	0.5
…						
…						

Normalize all the kernel values by dividing each value in the table with sum across each t

KERNEL REGRESSIONS

Predicted (or smoothed) value of kernel at each value of t is just **weighted sum of all the points at that value of t,** with the weights being the normalized kernel value.

This is our data table for the values of p across each t

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
p	16	14	15	20	17	15	12	18	21	24	26	24	20	28	30	31	31.5	26	32	35	29	30	31	32	28	34	32	29	30	29

Predicted value of our smoothed curve at t = 1 (for e.g.) is:

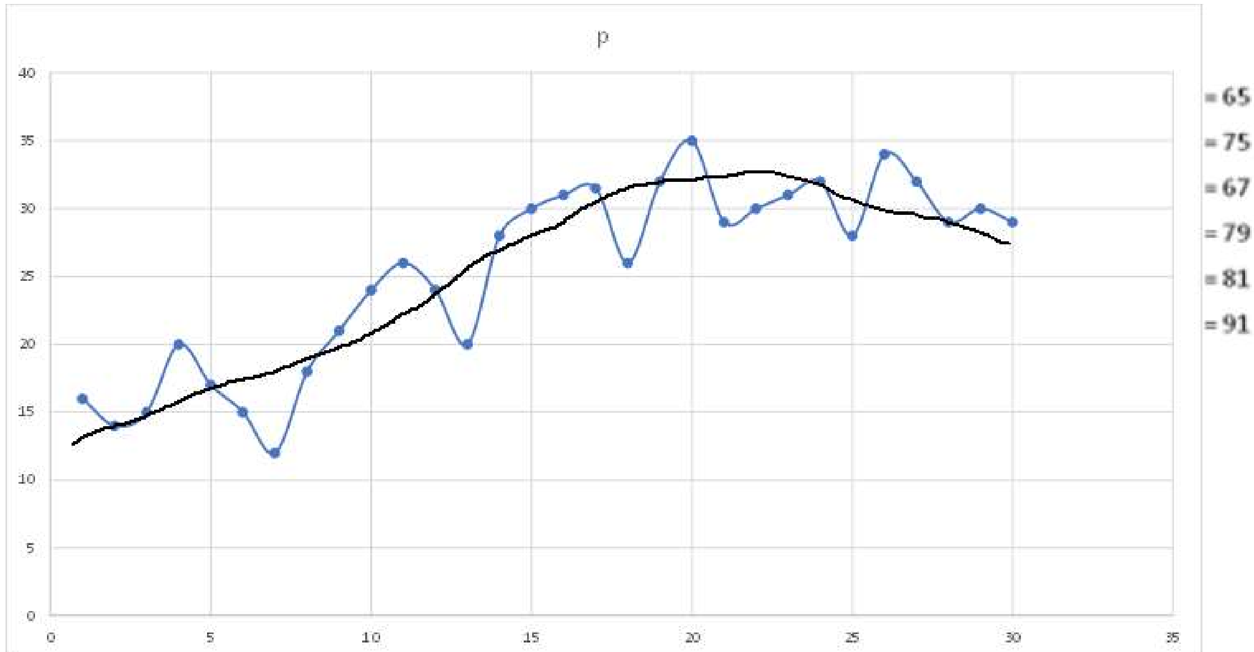
- (Value of p1 * normalized kernel value of p1 at t=1) +
- (Value of p2 * normalized kernel value of p2 at t=1) +
- (Value of p3 * normalized kernel value of p3 at t=1) +
- …
- (Value of p30 * normalized kernel value of p30 at t=1) +

KERNEL REGRESSIONS

This would be (substituting “16” for “value of p1”, “14” for “value of p2”, etc)

16*(0.08/0.3) + 14*(0.06/0.3) + 15*(0.05/0.3) + …

Ultimately, our ‘smoothed’ curve might look like this:



Generally, with a larger bandwidth parameter value, the curve will be ‘smoothed’ to a greater extent

BACKTESTING CHART PATTERNS

We begin by fitting kernel regressions to our sample of prices, P1, P2, …Pt

We fit one kernel regression for each rolling window. This implies that chart pattern should be completed within span of the window (after also allowing for some delay to avoid lookahead bias). In the paper, window length is set at 35 trading days

Using kernel representation of the price series, **we look for local extrema by finding times t such that the derivative of the kernel function changes sign around that time stamp.** This requires that we first express each chart pattern as a sequence of extrema

EXPRESSING CHART PATTERNS AS SEQUENCE OF EXTREMA

Definition 1 (Head-and-Shoulders) *Head-and-shoulders (HS) and inverted head-and-shoulders (IHS) patterns are characterized by a sequence of five consecutive local extrema $E_1, \ldots, E_5$ such that:*

$$HS \equiv \begin{cases} E_1 \text{ a maximum} \\ E_3 > E_1 \ , \ E_3 > E_5 \\ E_1 \text{ and } E_5 \text{ within } 1.5\text{percent of their average} \\ E_2 \text{ and } E_4 \text{ within } 1.5\text{percent of their average} \end{cases}$$

$$IHS \equiv \begin{cases} E_1 \text{ a minimum} \\ E_3 < E_1 \ , \ E_3 < E_5 \\ E_1 \text{ and } E_5 \text{ within } 1.5\text{percent of their average} \\ E_2 \text{ and } E_4 \text{ within } 1.5\text{percent of their average} \end{cases}$$

EXPRESSING CHART PATTERNS AS SEQUENCE OF EXTREMA

Definition 2 (Broadening) *Broadening tops (BTOP) and bottoms (BBOT) are characterized by a sequence of five consecutive local extrema $E_1, \ldots, E_5$ such that:*

$$BTOP \equiv \begin{cases} E_1 \text{ a maximum} \\ E_1 < E_3 < E_5 \\ E_2 > E_4 \end{cases} \quad , \quad BBOT \equiv \begin{cases} E_1 \text{ a minimum} \\ E_1 > E_3 > E_5 \\ E_2 < E_4 \end{cases}$$

Definitions for triangle and rectangle patterns follow naturally:

Definition 3 (Triangle) *Triangle tops (TTOP) and bottoms (TBOT) are characterized by a sequence of five consecutive local extrema $E_1, \ldots, E_5$ such that:*

$$TTOP \equiv \begin{cases} E_1 \text{ a maximum} \\ E_1 > E_3 > E_5 \\ E_2 < E_4 \end{cases} \quad , \quad TBOT \equiv \begin{cases} E_1 \text{ a minimum} \\ E_1 < E_3 < E_5 \\ E_2 > E_4 \end{cases}$$

EXPRESSING CHART PATTERNS AS SEQUENCE OF EXTREMA

Definition 4 (Rectangle) *Rectangle tops (RTOP) and bottoms (RBOT) are characterized by a sequence of five consecutive local extrema* $E_1, \ldots, E_5$ *such that:*

$$RTOP \equiv \begin{cases} E_1 \text{ a maximum} \\ \text{tops within 0.75 percent of their average} \\ \text{bottoms within 0.75 percent of their average} \\ \text{lowest top} > \text{highest bottom} \end{cases}$$

$$RBOT \equiv \begin{cases} E_1 \text{ a minimum} \\ \text{tops within 0.75 percent of their average} \\ \text{bottoms within 0.75 percent of their average} \\ \text{lowest top} > \text{highest bottom} \end{cases}$$

EXPRESSING CHART PATTERNS AS SEQUENCE OF EXTREMA

Definition 5 (Double Top and Bottom) *Double tops (DTOP) and bottoms (DBOT) are characterized by an initial local extremum E_1 and a subsequent local extrema E_a and E_b such that:*

$$
\begin{aligned}
E_a &\equiv \sup \{ P_{t_k^*} : t_k^* > t_1^* \ , \ k = 2, \ldots, n \} \\
E_b &\equiv \inf \{ P_{t_k^*} : t_k^* > t_1^* \ , \ k = 2, \ldots, n \}
\end{aligned}
$$

and

$$
DTOP \equiv \begin{cases} E_1 \text{ a maximum} \\ E_1 \text{ and } E_a \text{ within 1.5 percent of their average} \\ t_a^* - t_1^* > 22 \end{cases}
$$

$$
DBOT \equiv \begin{cases} E_1 \text{ a minimum} \\ E_1 \text{ and } E_b \text{ within 1.5 percent of their average} \\ t_a^* - t_1^* > 22 \end{cases}
$$

BACKTESTING CHART PATTERNS

With our **kernel representation of the price series in the rolling window**, we can detect extrema in this rolling window

Specifically, if m(t) is the kernel function, then:

- Maxima if $m'(t) > 0$ and $m'(t+1) < 0$ where $m'(.)$ is the first derivative of m(t) with respect to t
- Minima if $m'(t) < 0$ and $m'(t+1) > 0$ where $m'(.)$ is the first derivative of m(t) with respect to t

Once t corresponding to extrema in the rolling window has been identified, we proceed to identify maxima or minima in the original price series within [t-1, t+1]. This is used to determine whether a chart pattern has occurred according to the definitions in the previous slides

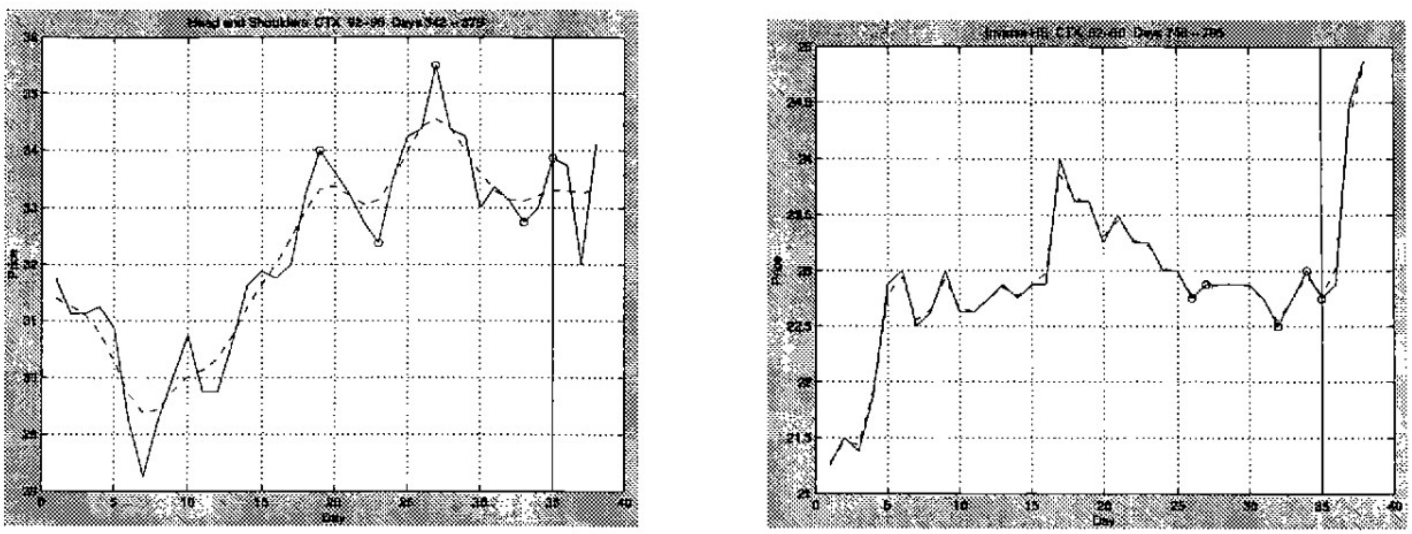
BACKTESTING CHART PATTERNS

One advantage of using the kernel regression with a smoothing parameter on bandwidth is that it ignores extrema that are “too small”.

Compare the alternative if we were to use the raw price data directly. i.e. we identify a price Pt as a local maxima if $P_{t-1} < P_t$ and $P_t > P_{t+1}$ (vice versa for a minima)

- This identifies too many extrema in the rolling window, and extrema which may not correspond to ‘major moves’ in the stock price
- The smoothing resulting from use of the kernel regression (larger bandwidth increases smoothing) will ensure that only extrema with significant amplitude are caught

KERNEL ESTIMATOR SMOOTHED VERSUS RAW



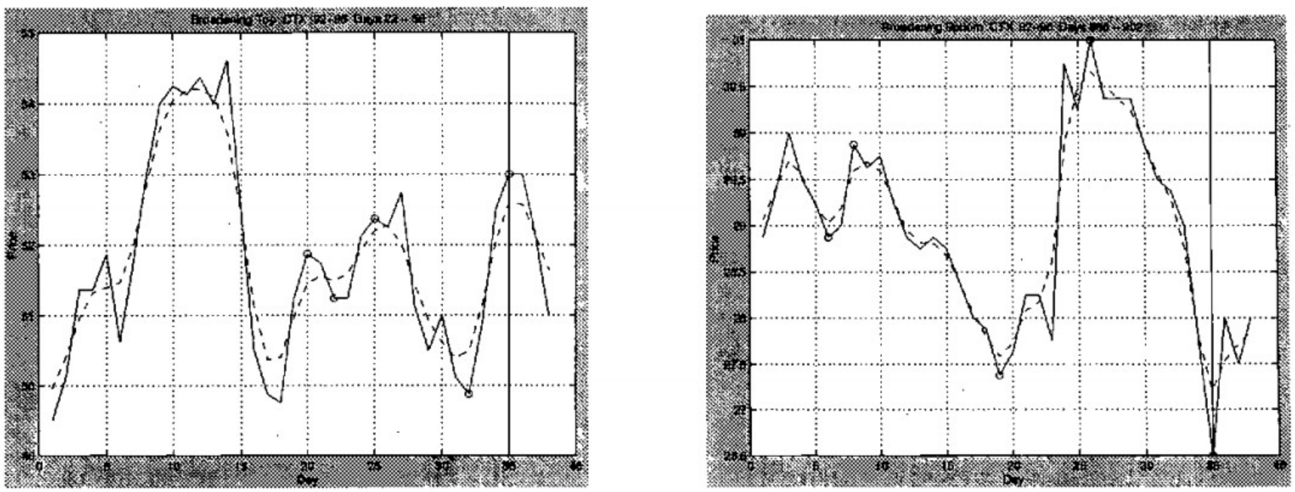
(a) Head-and-Shoulders

(b) Inverse Head-and-Shoulders

Figure III. Head-and-shoulders and inverse head-and-shoulders.

Source: https://www.nber.org/papers/w7613.pdf

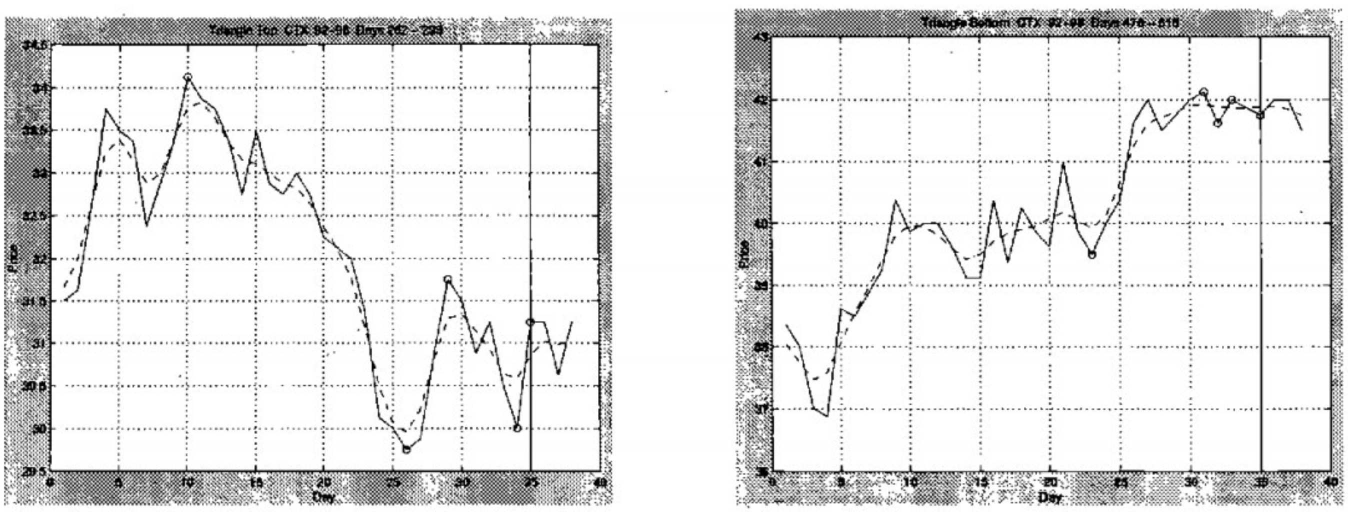
KERNEL ESTIMATOR SMOOTHED VERSUS RAW



(a) Broadening Top (b) Broadening Bottom

Figure IV. Broadening tops and bottoms.

KERNEL ESTIMATOR SMOOTHED VERSUS RAW



(a) Triangle Top (b) Triangle Bottom

Figure V. Triangle tops and bottoms.

MEASURING INFORMATIONAL CONTENT

The authors measure if the returns distribution conditional on existence of any one of the detected patterns is different from the unconditional returns distribution

In the null hypothesis where there is no additional information in the patterns, then there the conditional and unconditional distribution of returns should be similar

GOODNESS OF FIT TESTS

One simple approach is to compare the quantiles of conditional returns with their unconditional counterparts

If conditioning on technical patterns provide no incremental information, then quantiles of conditional returns should be similar to those of unconditional returns

This is done in 2 steps:

- First, we compute deciles of unconditional returns, and what breakpoint for those deciles are in our rolling window
- Second, we count the number of observations of conditional returns which fall into each of the deciles of unconditional returns

GOODNESS OF FIT TESTS

By definition, 10% of the unconditional distribution falls into each decile formed over the unconditional distribution

We compute the relative proportion of the conditional distribution that falls into each of the deciles formed above, and test to see if it is significantly different from 10%

In the conditional case, returns are defined as the 1-day return d days after the pattern has concluded. Each set of returns are grouped by the pattern that caused it. Since the authors have 10 different patterns, they end up with 10 sets of conditional returns

In the unconditional case, returns are computed across the entire sample of stock prices

COMPARISON STATISTICS

Table IV

Summary statistics (mean, standard deviation, skewness, and excess kurtosis) of raw and conditional 1-day normalized returns of NASDAQ stocks from 1962 to 1996, in 5-year subperiods, and in size quintiles. Conditional returns are defined as the daily return three days following the conclusion of an occurrence of one of 10 technical indicators: head-and-shoulders (HS), inverted head-and-shoulders (IHS), broadening top (BTOP), broadening bottom (BBOT), triangle top (TTOP), triangle bottom (TBOT), rectangle top (RTOP), rectangle bottom (RBOT), double top (DTOP), and double bottom (DBOT). All returns have been normalized by subtraction of their means and division by their standard deviations.

Moment	Raw	HS	IHS	BTOP	BBOT	TTOP	TBOT	RTOP	RBOT	DTOP	DBOT
					All Stocks, 1962 to 1996						
Mean	0.000	−0.016	0.042	−0.009	0.009	−0.020	0.017	0.052	0.043	0.003	−0.035
S.D.	1.000	0.907	0.994	0.960	0.995	0.984	0.932	0.948	0.929	0.933	0.880
Skew	0.608	−0.017	1.290	0.397	0.586	0.895	0.716	0.710	0.755	0.405	−0.104
Kurt	12.728	3.039	8.774	3.246	2.783	6.692	3.844	5.173	4.368	4.150	2.052

Note these are normalized statistics

Source: https://www.nber.org/papers/w7613.pdf

HOW DO THESE COMPARE WITH EX ANTE EXPECTATIONS?

	Normalized mean	Ex ante expectations	Consistent?
Head and Shoulders	-0.016	Bearish	Yes
Inverse head shoulders	0.042	Bullish	Yes
Broadening tops	-0.009	Bearish	Yes
Broadening bottoms	0.009	Bullish	Yes
Triangle top	-0.02	Bearish	Yes
Triangle bottom	0.017	Bullish	Yes
Rectangle top	0.052	Bullish	Yes
Rectangle bottom	0.043	Bearish	No
Double top	0.003	Bearish	No
Double bottom	-0.035	Bullish	No

Score: 7/10

HOW DO THESE COMPARE WITH EX ANTE EXPECTATIONS?

	Normalized mean	Ex ante expectations	Consistent?
Head and Shoulders	-0.016	Bearish	Yes
Inverse head shoulders	0.042	Bullish	Yes
Broadening tops	-0.009	Bearish	Yes
Broadening bottoms	0.009	Bullish	Yes
Triangle top	-0.02	Bearish	Yes
Triangle bottom	0.017	Bullish	Yes
Rectangle top	0.052	Bullish	Yes
Rectangle bottom	0.043	Bearish	No
Double top	0.003	Bearish	No
Double bottom	-0.035	Bullish	No

We interpret the economic value of these statistics using row (1), "Head and Shoulders" as example:
"The 1 day return of an asset upon completion of a head and shoulders pattern 1 day ago is 0.016 standard deviations below that of its usual (mean) return"

IS THERE A FUNDAMENTAL ECONOMIC EXPLANATION?

- *Even if* the chart patterns overall work, what is the reason for this?
- Behavioural economics discusses how psychological factors and biases driving market participants may (in the aggregate, on a large scale) explain price movements
- For instance, investor anchoring and underreaction may explain the existence of long-term price trends caused by news rather than a sharp one time adjustment
- Investor psychology may also explain the existence of specific price levels (“support” and “resistance” levels) that underpin the patterns discussed
- We will discuss behavioural economics as a possible economic reason behind some of these chart patterns in class 4

FAST FORWARD TO 2024: SUPPORT VECTOR MACHINES FOR PATTERN RECOGNITION

1. In 2024, we have **additional tools to handle pattern recognition** [e.g. chart patterns] in financial time series.

2. 'Benchmark' methodology based on existing technology for pattern recognition in financial time series is support vector machines. **This has superseded kernel regressions due to better performance**

3. In this section, we will **overview support vector machines,** and also discuss basic implementation in Python

Support Vector Machines (SVMs) for TS classification

Support Vector Machines (SVM)

- **Support Vector Machines (SVMs)** are a widely used technique in machine learning for classification as of 2024
- It is formally a **supervised learning technique** that requires training data, and is used for classification
- SVMs are **usually applied on extracted features**, rather than raw datasets

Features of time series

- A feature is a transformation of the data, or part of the data, to highlight information that is relevant to economic theory, or to the structure of the problem
- In many cases, time series classification proceeds based on "features" extracted from time series.
- Feature extraction is an intermediate step in many machine learning algorithms, where the ML algorithms work on the features rather than the raw data directly. Prior domain specific knowledge of the economic or underlying structure of the data can be useful for feature identification and extraction
- Examples of features:
  - "Volatility", "Interquartile range" of time series
  - "Frequency" of cycles in the time series
  - "N period moving average" of time series
  - "AR order" of the time series
  - Average time between missing values
- To compute distance between time series based on their features, we commonly compute a Euclidean distance measure over differences in the feature values (not the raw time series values). Other measures (e.g. 1-norm) might also be used

Types of features for time series data

- **Summary statistics**
  - Mean, variance, drift
- **Model based**
  - ACF values, PACF values (with or without differences)
  - Seasonal patters
  - Other attributes of time series: spectral entropy, hurst coefficient (we will discuss in next few slides – for now just think of these as properties of the time series)
- **Time series metadata** ("data about data")
  - Frequency, number of missing observations, average time between missing observation, originating sensor, time zone

(list is not exhaustive)

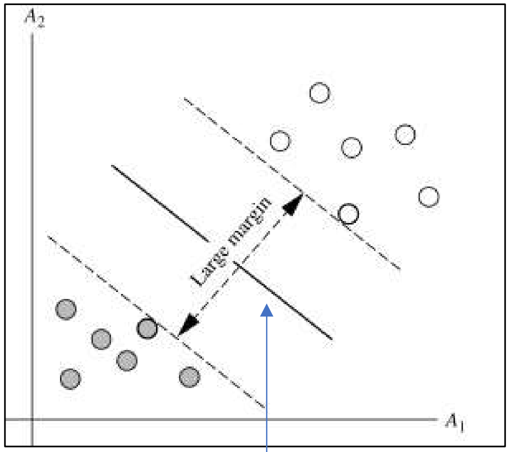
SVM terminology

- SVM separates datapoints into groups by searching for the **location of a separating line or plane between both sets of datapoints**

- The separating line or plane is formally called a "**separating hyperplane**" to allow for high dimensionality feature set

- The algorithm locates separating hyperplane to **maximize its distance from the nearest point** in each of the two sets that it is separating

Source: https://www.sciencedirect.com/topics/computer-science/separating-hyperplane

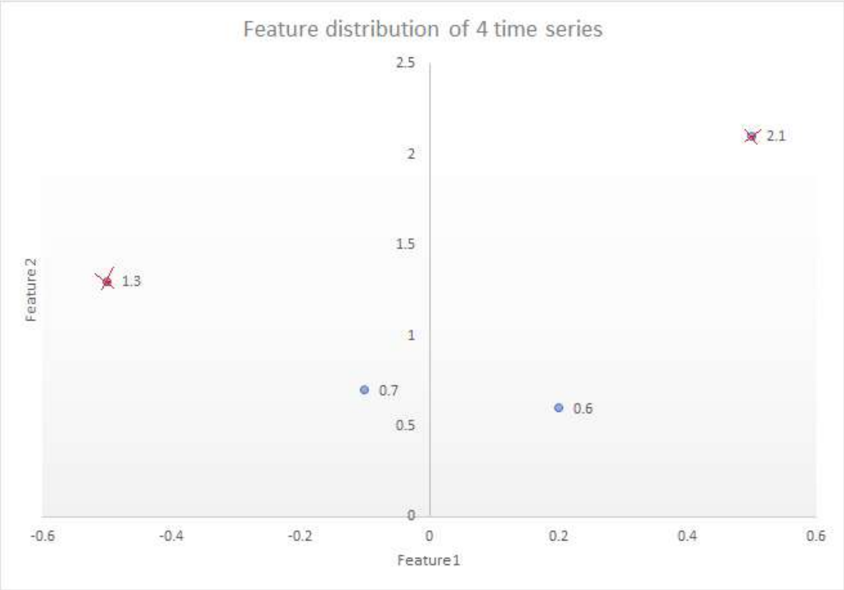


"**Separating Hyperplane**". For a 2 dimensional feature set, this is nothing more than a simple line

Introduction to SVMs

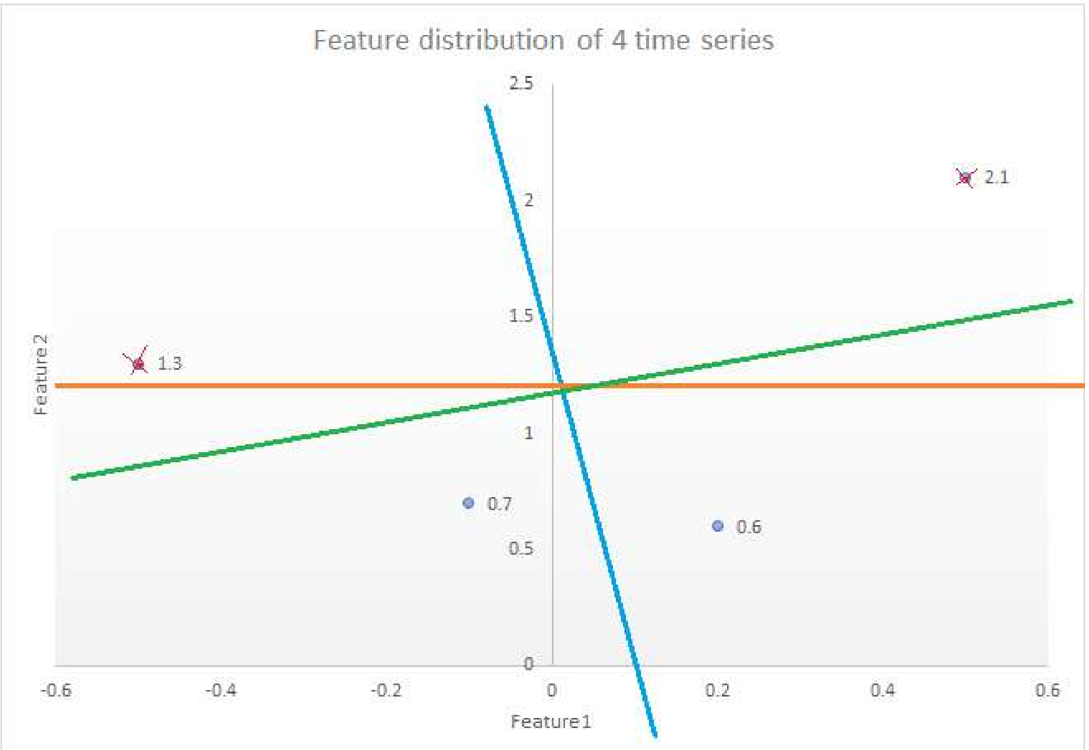
- Assume we have 4 time series, and 2 features computed as follows

- We have hand labelled (or divided) time series into 2 groups, to form training set



Variable	Feature 1 (Mean)	Feature 2 (Variance)	Group
Time series 1	0.5	2.1	1
Time series 2	-0.5	1.3	1
Time series 3	0.2	0.6	2
Time series 4	-0.1	0.7	2

Introduction to SVMs

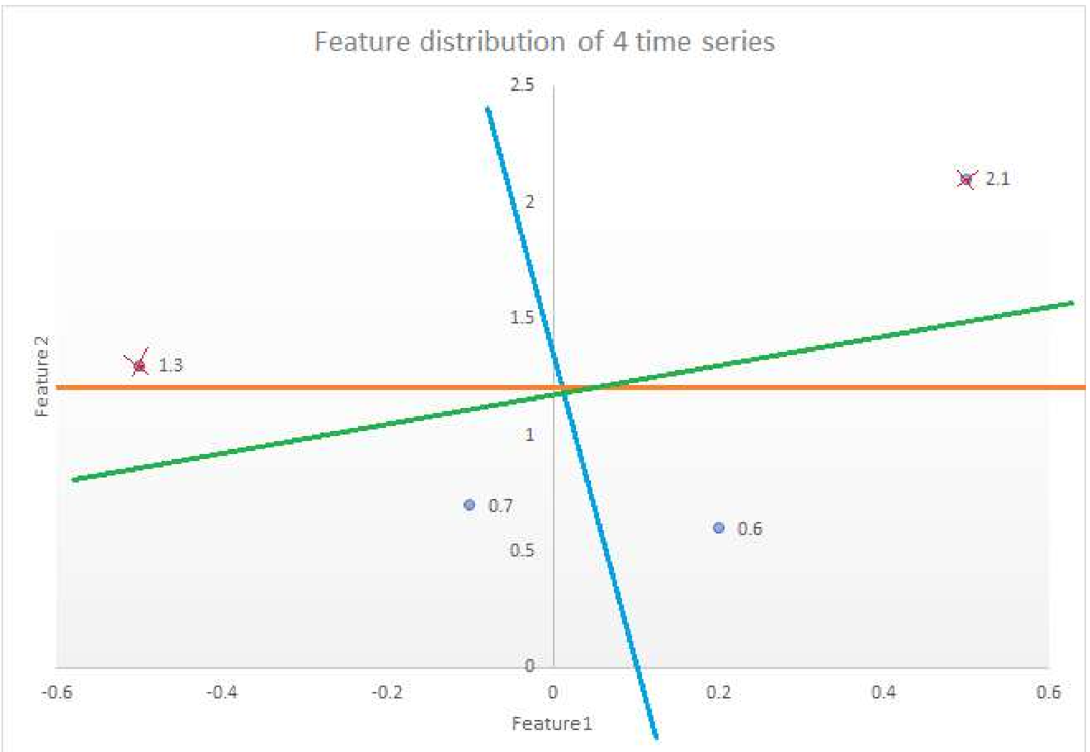


1. Superimposed are 3 possible options for separating hyperplanes to separate data into 2 groups.

2. Feature set is 2 dimensional (i.e. 2 features). Hence separating hyperplane is just a line for this example

3. There are an infinite number of such possible separating lines

Introduction to SVMs

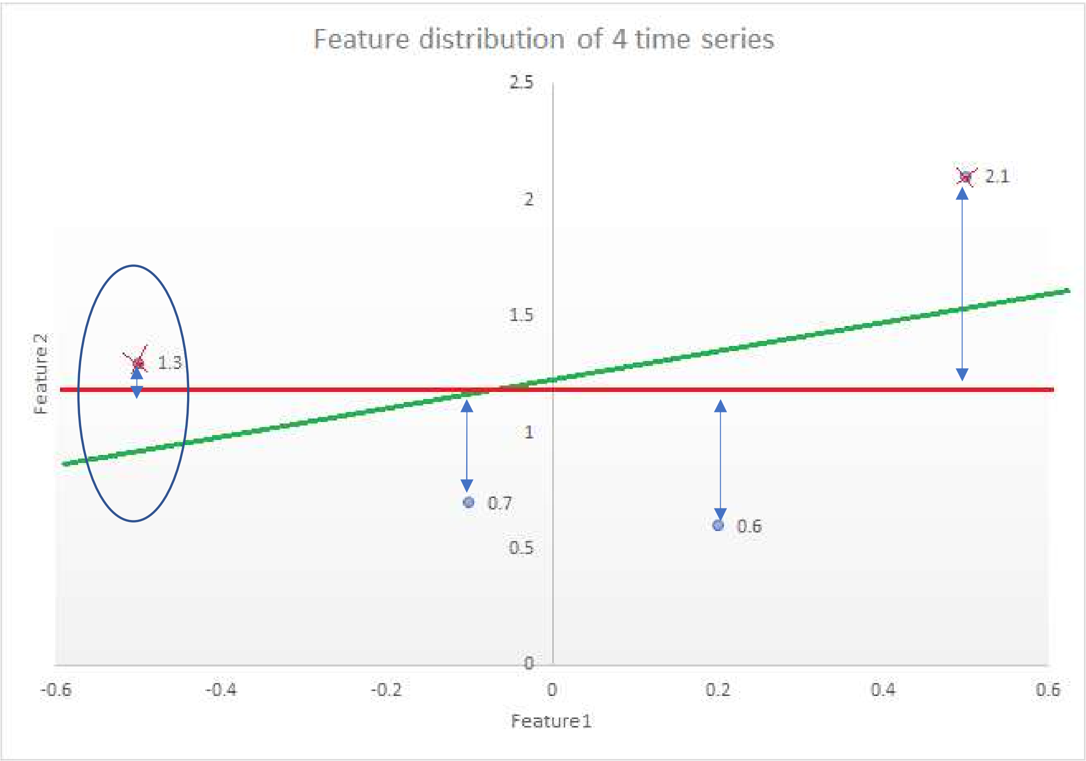


1. Blue line is not a viable option – why?

2. Red line indeed separates data into 2 groups (as defined in training set).

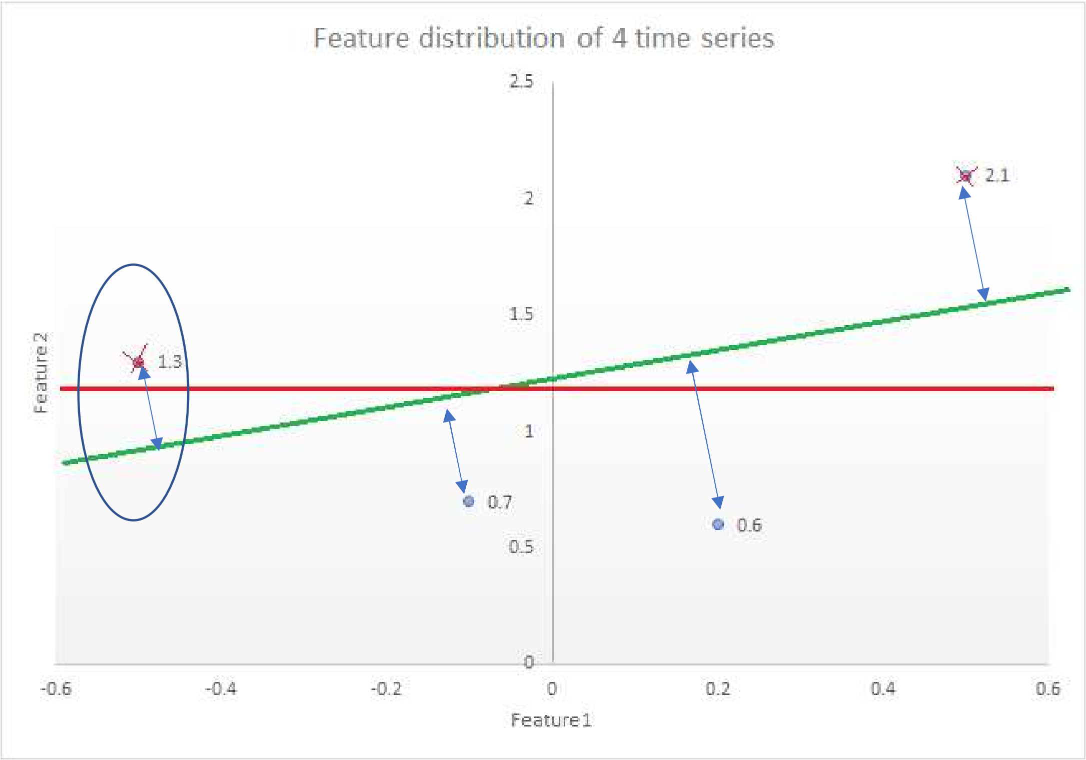
3. Green line also separates data into 2 groups (as in training set), and also has a higher minimum distance between the closest point from each set

Introduction to SVMs



1. Blue line removed for clarity
2. Blue color “double headed arrows” show distance between red line and points from both sets
3. “Margin” of the red line is the distance highlighted with the oval (around point with y = 1.3). This is the smallest distance between the red line and the points from both groups

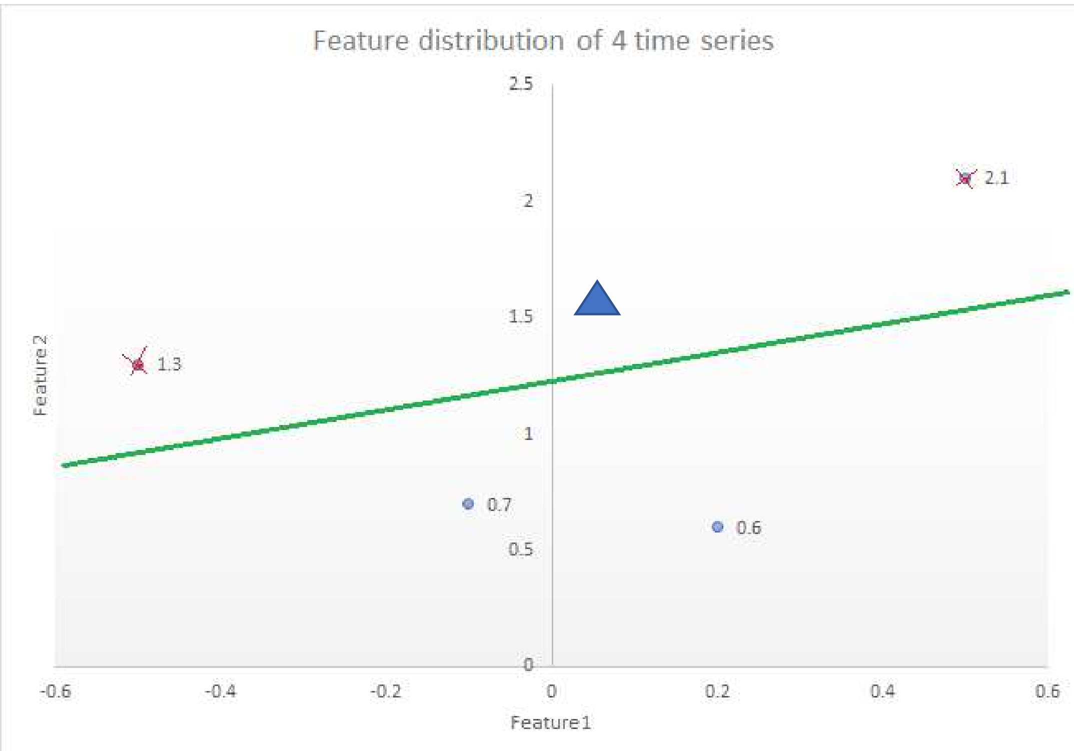
Introduction to SVMs



1. "Margin" of the green line is also the distance highlighted with the oval (around point with y = 1.3). **This is smallest distance between green line and the points from both groups**

2. As the green line has a larger margin than the red line, **we pick the green line as the hyperplane that separates both groups in the training set**

Introduction to SVMs



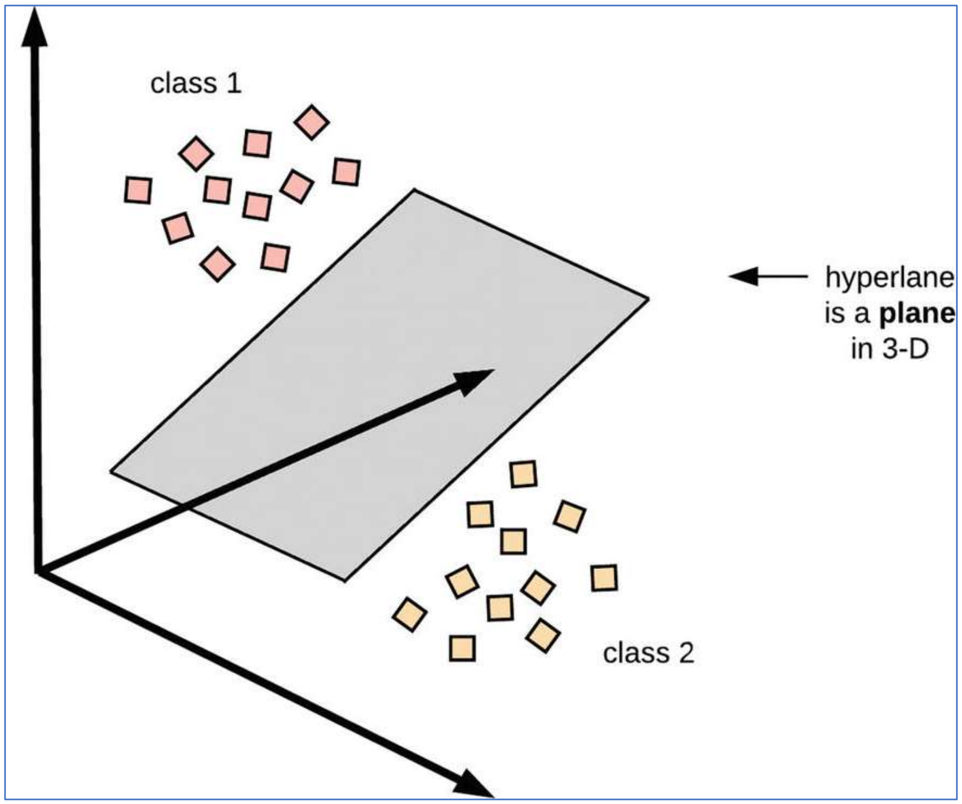
1. “Output” of SVM is **therefore equation of separating hyperplane**. In this case just equation of green line

2. This is called the **maximum margin classifier**

3. How do we use this result?
    - If given a new “unknown” time series to classify (e.g., blue rectangle in diagram), we check to see **which side of separating hyperplane new time series falls on**.

    - This **decides its category**

Introduction to SVMs



Source: https://link.springer.com/chapter/10.1007/978-1-4842-4470-8_22

1. Extending the reasoning, if we had a feature set with 3 features (instead of just 2), then the separating **hyperplane will be a plane, instead of just a line**

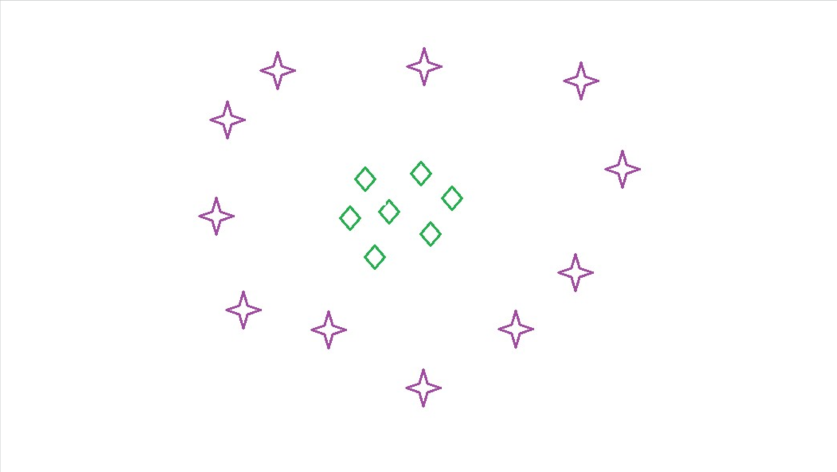
Variable	Feature 1 (Mean)	Feature 2 (Variance)	Feature 3 (% change)	Group
Time series 1	0.5	2.1	6%	1
Time series 2	-0.5	1.3	5%	1
Time series 3	0.2	0.6	1%	2
Time series 4	-0.1	0.7	2%	2
….				
Time series N	…	…	…	…

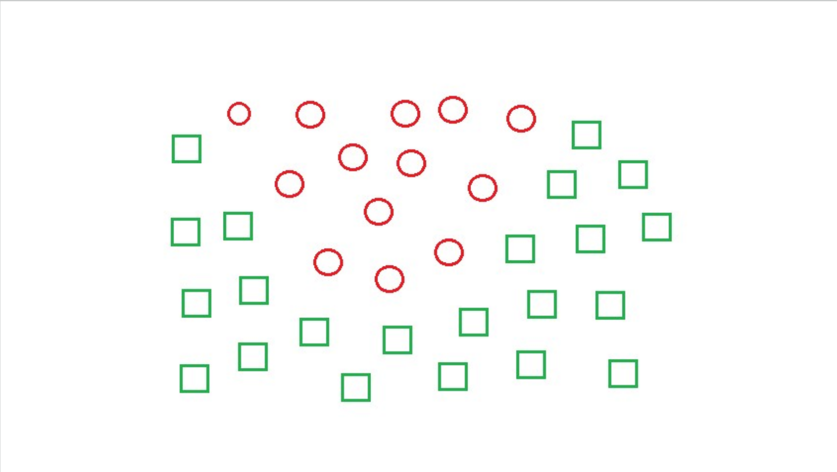
SVMs: Non Separable Data

- What if training set (with just 2 features) looks like this? Can we still use SVM to classify data?

- How can you draw a separating hyperplane (i.e. a straight line) through the data points to separate them into both groups?

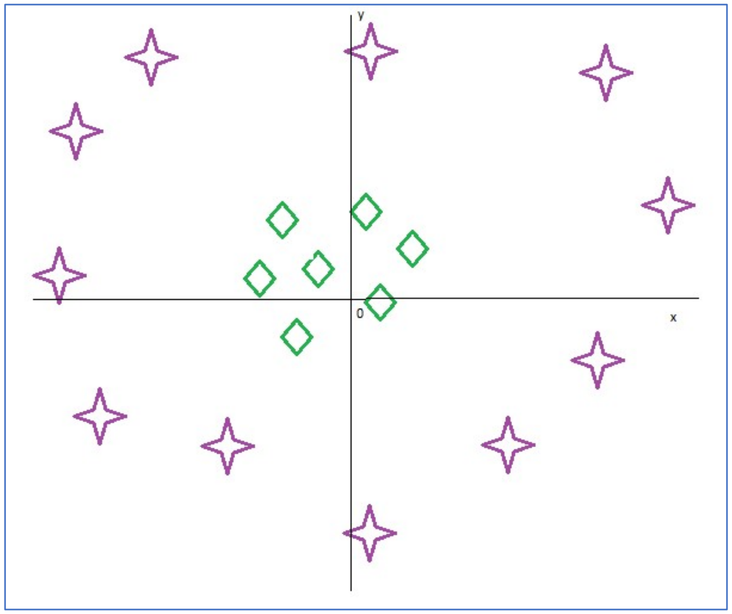
- Answer:
  - This cannot be done in 2 dimensions
  - However, even though feature set is restricted to 2 dimensions, we are not

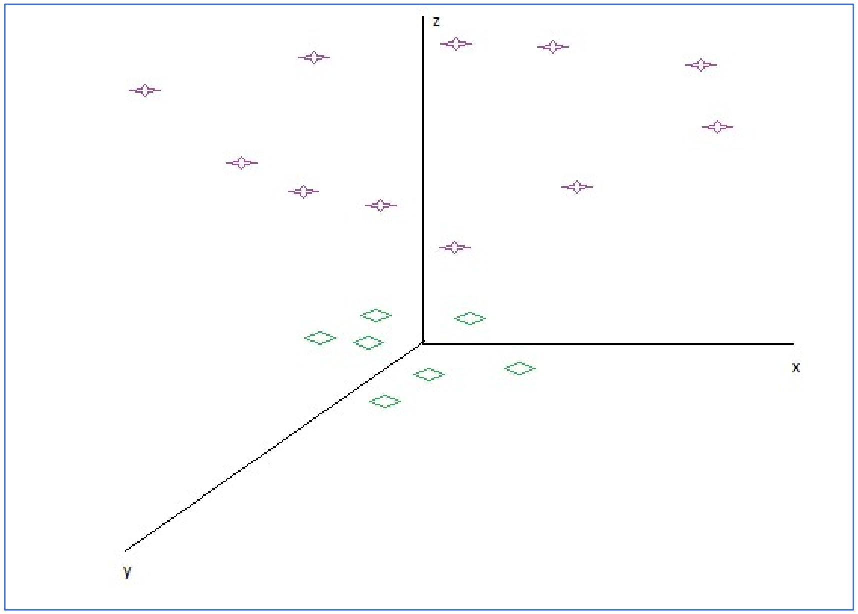




SVMs: Non Separable Data

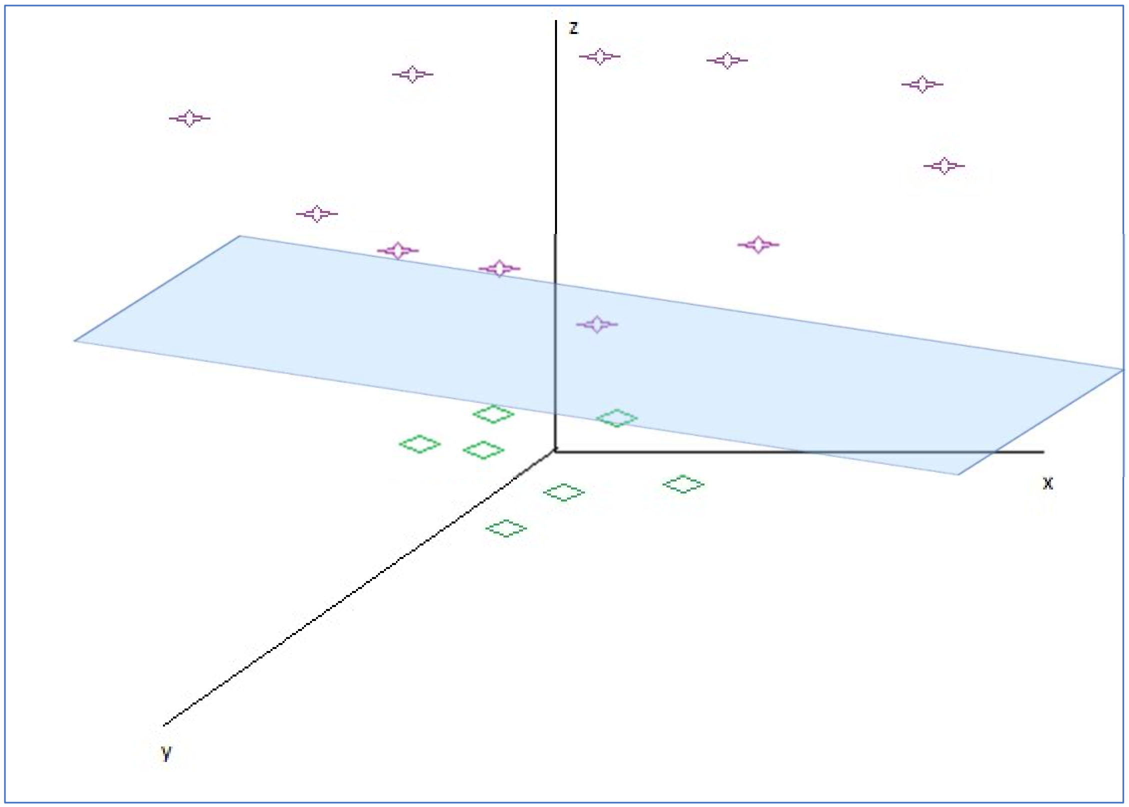
- Consider a transformation **$z = x^2 + y^2$ (i.e. equation of a circle)**
- We compute this **transformation for each object** (i.e. time series) in our training set, and plot revised data set in 3 dimensions (x,y,z axis)
- New coordinate of each object (x,y,z) is its **original x and y coordinate, plus its new z coordinate**



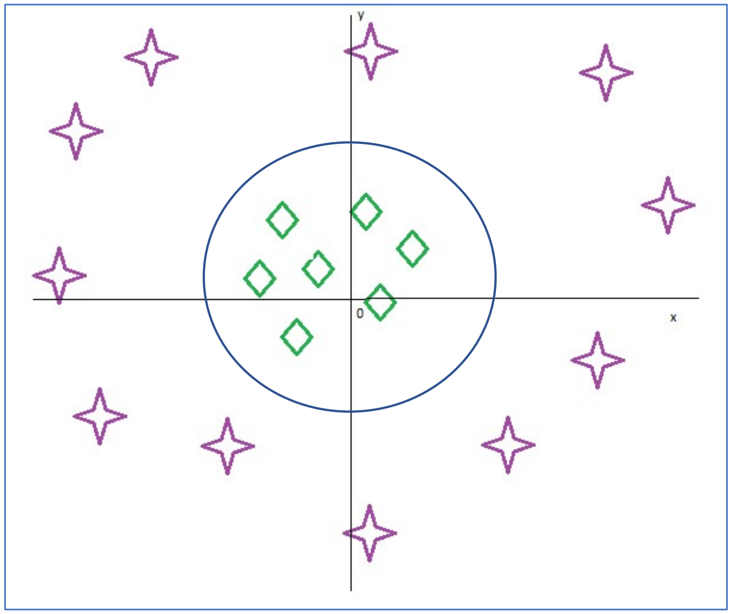


SVMs: Non Separable Data

- Working with 3 dimensional data, can we now hypothesize a separating hyperplane for this training dataset?

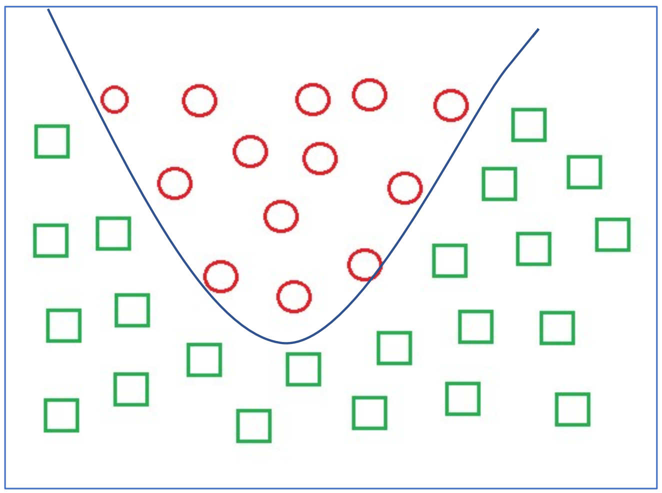


SVMs: Non Separable Data



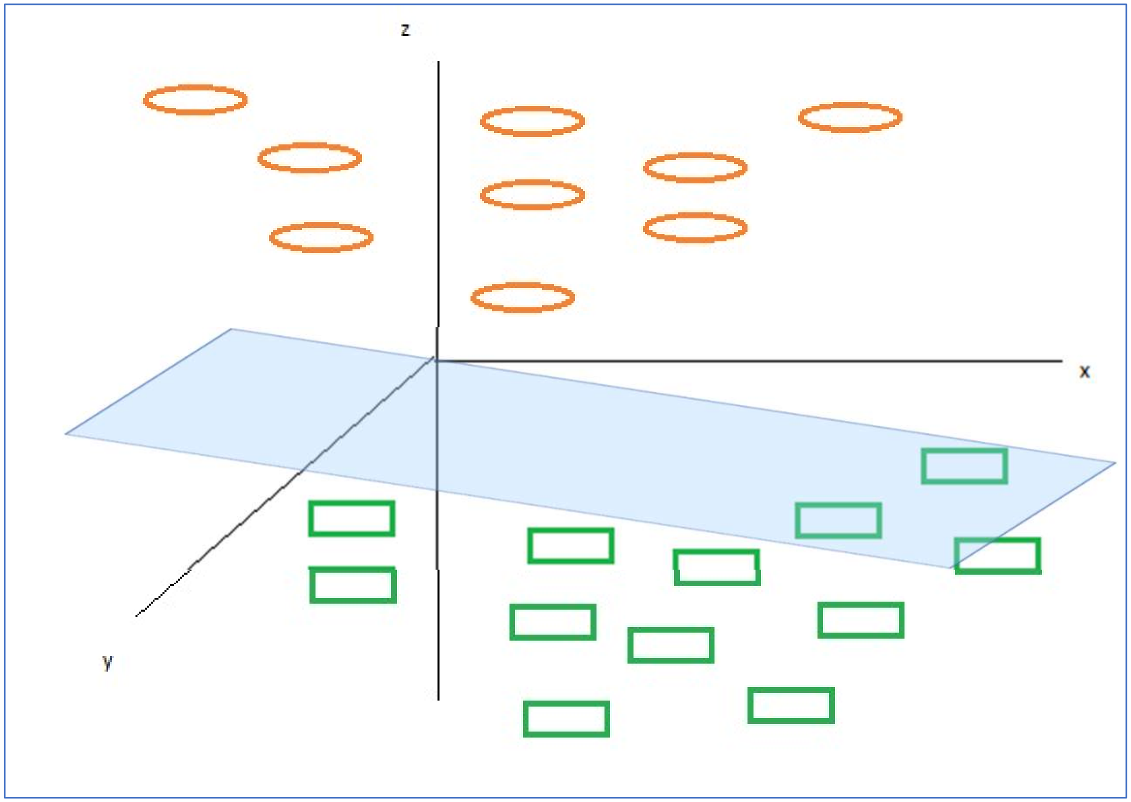
- Recall $z = x^2 + y^2$ was **equation of a circle**
- What does **new separation look like** in our original data?
- Note: **SVM algorithm still needs to compute an appropriate radius and center for this circle** to fully define its location as a maximum margin classifier
- Overall idea: when we define a new dimension to the data, we wish to shift all objects in the same group in **the same direction** “within” the new dimension, and all objects in a different group in **opposite direction**

SVMs: Non-Separable Data



- How about this?
- We can draw parabola through data to separate it into two groups. i.e. $y = x^2$
- Consider transformation $z = y - x^2$ (note: there is probably more than 1 way to do this. Some constants, lower order terms may be needed based on precise location of points)
- Points above the parabola have $y > x^2$ (i.e. $y - x^2 > 0$)
- Points below the parabola have $y < x^2$ (i.e. $y - x^2 < 0$)
- Location of separating hyperplane is $z = 0$? (with adjustments to obtain maximum margin classifier)

SVMs: Non-Separable Data



- Location of separating hyperplane in light blue (e.g. z = 0)

Discussion on class project

DISCUSSION ON CLASS PROJECT

The class project should be presented by all groups towards end of this term

In this project, you should, in your groups, implement a quantitative trading strategy.

There are no limitations on asset class or trading styles. The trading strategy should however be fully systematic

DISCUSSION ON CLASS PROJECT

The class project should result in following output:

1. Written project proposal (say 1 to 2 pages) to be submitted via eLearn around class 5. This will allow us to provide early feedback

2. Approximately 20 minute group presentation to be conducted during regular class session in the designated week in course outline towards end of term

3. Written report to be submitted via eLearn before start of recess week

DISCUSSION ON CLASS PROJECT

It is recommended to consider following points in execution of project:

- Consider and describe what are fundamental economic intuition and forces behind this strategy working. E.g. in case of pairs trading, economic intuition is "both legs of a pair share similar underlying economics", and for some patterns surveyed in today's class, it may be due to behavioural reasons. We will discuss more strategies over next few classes and their underlying motivations

- Logistics of implementing this strategy: where will you get the data from, what econometric or computational techniques will you use, which trading universe (and why?), what trading frequency, etc

- Be able to show a backtested strategy result in your presentation and final report

- Realism of backtest: ability to trade instruments under consideration in both long and short directions, trading costs, liquidity, etc

- Discussion of backtested performance, strengths / weaknesses, areas for further research

FOR OUR NEXT CLASS ..

In the next class, we will:

1. Overview Python code for using SVMs to detect and backtest chart patterns
2. Examine in more detail macro economic, micro economic and corporate finance drivers behind selected statistical arbitrage strategies
3. We will discuss broadly the state of existing academic research with hundreds of cross sectional market anomalies, or 'factors', before deep diving into the construction of selected strategies to examine their methodology
4. Fundamental economic effects discussed will include managerial moral hazard, incomplete contracting and search costs
5. We will discuss how the micro economic manifestation of these forces at an individual corporate level may result in large scale properties across thousands of equity instruments which can be identified statistically