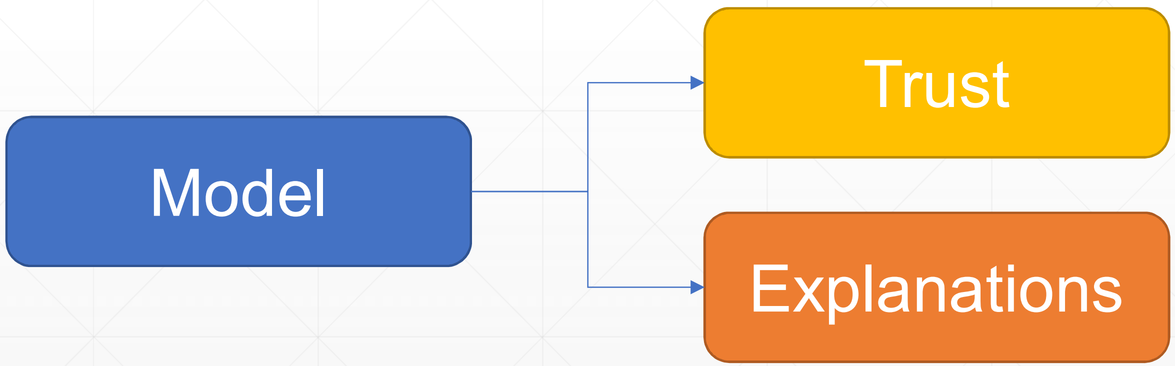
以LIME模型和關聯法則分析深度學習優化交易策略

指導教授： 中山財管所 王昭文 教授
簡報製作： 中山財管所 蘇彥庭 研究助理

LIME模型介紹

- 透過LIME模型(Ribeiro et al., 2016)，解決機器學習模型分類器為黑盒子(Black-boxes)的問題，讓使用者有理由相信機器學習模型的決策過程是可信的，並且能夠解釋模型為何會預測這樣的結果。

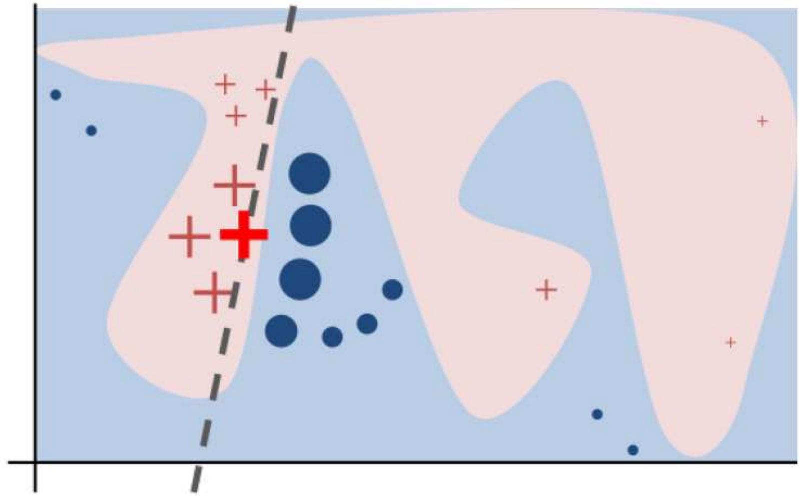
If the users do not trust a model or a prediction, they will not use it.



LIME模型介紹

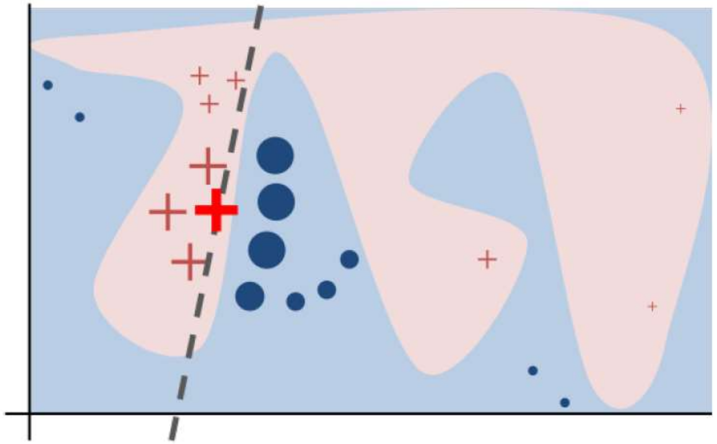
- 英文模型名稱：Local Interpretable Model-Agnostic Explanations (LIME)
- 中文模型名稱：局部可解讀的與模型無關之解釋
- 模型能解決問題：
  - Choose between competing models.
  - Detect and improve untrustworthy models.
  - Get insights into the model.
- 模型概念：
  - 在機器學習已訓練好的分類器模型下，挑選代表性樣本，並透過人類可理解的擾亂輸入特徵方式，以一個簡單的的模型來局部逼近此複雜模型，觀察模型的預測結果變化，藉以推論模型如何進行預測。

LIME模型方法說明



- G：使用者可解釋的模型集合
- g：可解釋的模型， $g \in G$
- f：原始模型
- $f(x)$：樣本點x在原始模型上對於某分類的預測機率
- $\pi_x(z) = \exp(-\frac{D(x,z)^2}{\sigma^2})$：樣本點$z$與樣本點$x$的距離，樣本加權值(愈接近樣本點$x$權值愈大)， D為距離函數，σ為寬度參數
- $\mathcal{L}(f, g, \pi_x)$：損失函數，表示在π_x上，以原始的f模型預測與可解釋的g模型預測的差距
- $\Omega(g)$：模型複雜度(例如決策樹模型即為深度、線性迴歸模型則為係數項非零的變數個數)
- LIME模型的目標： $\xi(x) = \underset{g \in G}{\operatorname{argmin}} \mathcal{L}(f, g, \pi_x) + \Omega(g)$

LIME模型方法說明



- 局部抽樣方法
  - 選擇一個樣本點x ，將此樣本點轉換到可解釋的空間$x'\epsilon\{0,1\}^{d'}$ (圖像)
  - 以均勻分配方式隨機抽樣n個樣本點z'
  - 將可解釋空間的每個樣本點z'還原到原始空間樣本點z，並且以原始模型f進行預測得出預測值
- LIME模型的目標： $\xi(x) = \underset{g \in G}{\operatorname{argmin}} \mathcal{L}(f, g, \pi_x) + \Omega(g)$
- $\mathcal{L}(f, g, \pi_x) = \sum_{z, z' \in Z} \pi_x(z)(f(z) - g(z'))^2$
- $g(z') = \omega_g \cdot z'$

LIME模型

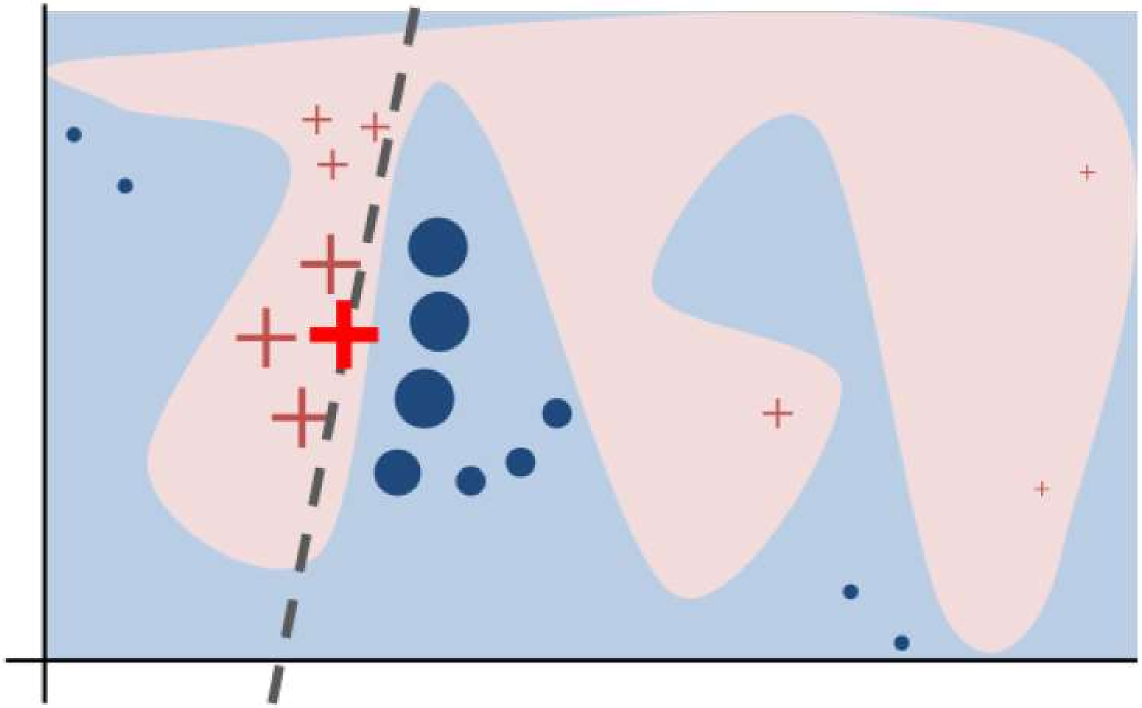


Figure 3: Toy example to present intuition for LIME. The black-box model's complex decision function f (unknown to LIME) is represented by the blue/pink background, which cannot be approximated well by a linear model. The bold red cross is the instance being explained. LIME samples instances, gets predictions using f, and weighs them by the proximity to the instance being explained (represented here by size). The dashed line is the learned explanation that is locally (but not globally) faithful.

關聯法則介紹

- 關聯法則(Association Rule)
- 常用於購物籃分析(Basket Data Analysis)，從每個客戶每次購物的清單中進行分析，找出客戶常購買的商品組合資訊
- 最常聽到的案例：沃爾瑪的啤酒與尿布的故事
- 關聯法則的演算法相當多種，此處以常見的Apriori 演算法進行實作，R的套件名稱為arules

關聯法則參數

- 支持度(Support) $= \dfrac{X和Y同時出現的次數}{所有交易次數}$

- 信心水準(Confidence) $= \dfrac{X和Y同時出現的次數}{X出現的次數}$

- 支持度與信心水準兩指標愈高，代表這個組合是很常出現的。

- 透過設定上述兩個參數的最小值，找出有用的關聯規則。

關聯法則範例

- 分析的關聯規則以 X->Y 表示， X稱為前項（antecedents，R軟體稱為lhs，即left hand sides），Y稱為後項（consequents，R軟體稱為rhs，即right hand sides）。我們舉以下的Transaction項目列表為例：

TID	*Items*
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diaper, Beer, Coke
4	Bread, Milk, Diaper, Beer
5	Bread, Milk, Diaper, Coke

- 以下是幾個規則範例：

{Milk, Diaper} a {Beer} (s=0.4, c=0.67)
{Milk, Beer}a{Diaper} (s=0.4, c=1.0)
{Diaper, Beer}a{Milk} (s=0.4, c=0.67)
{Beer}a{Milk, Diaper} (s=0.4, c=0.67)

- { X } a { Y } (支持度(s)， 信心水準(c))

此處範例參考： http://www.cc.ntu.edu.tw/chinese/epaper/0036/20160321_3606.html

以LIME模型分析深度學習交易策略

- LIME模型可用R的lime套件進行實作，但因為lime模型並未支援keras模型，所以需要自製函數讓lime能和keras模型進行結合

```
## 以LIME模型進行分析
# 建立函數讓Keras能和Lime做對接
model_type.keras.models.Sequential <- function(x, ...){
  return("classification")
}

predict_model.keras.models.Sequential <- function(x, newData, type, ...){
  pred <- predict_proba(object=x, x=as.matrix(newData))
  return(data.frame(Yes=pred[,2], No=pred[,1]))
}
```

- 需要注意的是，上面這兩個函數的名稱不可變換

以LIME模型分析深度學習交易策略

- 為了解深度學習是如何預測漲分類，以及找出潛在未知的特徵，此處指分析預測漲實際漲，且預測漲機率為55%以上之樣本

```
# 建立LIME模型解釋器
explainer <- lime(x=as_tibble(trainData), model=model, bin_continous=F)

# 選取預測漲實際漲且預測漲機率為0.55以上的樣本
predictCorrectIdx <- which(as.numeric(testRet>0)==predict & predict==1 & predictProb[,2]>=0.55)
predictCorrectTestRet <- testRet[predictCorrectIdx]
predictCorrectTestData <- testData[predictCorrectIdx,]
explainSample <- as_tibble(predictCorrectTestData)

# 模型解釋
explanation <- explain(x= explainSample,         # 解釋樣本
                       explainer= explainer,     # LIME模型解釋器
                       n_labels= 1,              # 解釋1個分類(漲)
                       n_features= 5,            # 擷取最高的5個特徵
                       kernel_width= 0.5)        # LIME模型解釋器參數，用於衡量樣本距離
```

以LIME模型分析深度學習交易策略

	model_type	case	label	label_prob	model_r2	model_intercept	model_prediction	feature	feature_value	feature_weight	feature_desc	data	prediction
1	classification	1	Yes	0.5657537	0.0000000000000000161841909	0.5657537	0.5657537	vol_5MA	0.340576887	-0.00000000000000000400373186	0.332 < vol_5MA <= 0.388	list(vol_5MA = 0.340576887260995, tradeVolume = 0.348...	list(Yes = 0.565753698348999, No = 0.434246271848679)
2	classification	1	Yes	0.5657537	0.0000000000000000161841909	0.5657537	0.5657537	weekMACD	0.863529598	-0.00000000000000002222927661	0.740 < weekMACD	list(vol_5MA = 0.340576887260995, tradeVolume = 0.348...	list(Yes = 0.565753698348999, No = 0.434246271848679)
3	classification	1	Yes	0.5657537	0.0000000000000000161841909	0.5657537	0.5657537	dayBBdn	0.949209125	0.00000000000000000776957456	0.569 < dayBBdn	list(vol_5MA = 0.340576887260995, tradeVolume = 0.348...	list(Yes = 0.565753698348999, No = 0.434246271848679)
4	classification	1	Yes	0.5657537	0.0000000000000000161841909	0.5657537	0.5657537	bottom4PricePer	0.387149811	-0.00000000000000000729999474	0.333 < bottom4PricePer <= 0.500	list(vol_5MA = 0.340576887260995, tradeVolume = 0.348...	list(Yes = 0.565753698348999, No = 0.434246271848679)
5	classification	1	Yes	0.5657537	0.0000000000000000161841909	0.5657537	0.5657537	prLeveldayMA90	0.162529031	-0.00000000000000000665563644	prLeveldayMA90 <= 0.258	list(vol_5MA = 0.340576887260995, tradeVolume = 0.348...	list(Yes = 0.565753698348999, No = 0.434246271848679)
6	classification	2	Yes	0.6129197	0.0000000000000000384324851	0.6129197	0.6129197	vol_5MA	0.349135571	0.00000000000000003209638941	0.332 < vol_5MA <= 0.388	list(vol_5MA = 0.349135571180205, tradeVolume = 0.345...	list(Yes = 0.612919688224792, No = 0.38708034157753)
7	classification	2	Yes	0.6129197	0.0000000000000000384324851	0.6129197	0.6129197	dayFastD	0.170128801	0.00000000000000000958821994	dayFastD <= 0.227	list(vol_5MA = 0.349135571180205, tradeVolume = 0.345...	list(Yes = 0.612919688224792, No = 0.38708034157753)
8	classification	2	Yes	0.6129197	0.0000000000000000384324851	0.6129197	0.6129197	dayBBup	0.396290185	0.00000000000000003123226647	0.382 < dayBBup <= 0.571	list(vol_5MA = 0.349135571180205, tradeVolume = 0.345...	list(Yes = 0.612919688224792, No = 0.38708034157753)
9	classification	2	Yes	0.6129197	0.0000000000000000384324851	0.6129197	0.6129197	monthBBup	0.362422140	0.00000000000000000638150277	0.290 < monthBBup <= 0.384	list(vol_5MA = 0.349135571180205, tradeVolume = 0.345...	list(Yes = 0.612919688224792, No = 0.38708034157753)
10	classification	2	Yes	0.6129197	0.0000000000000000384324851	0.6129197	0.6129197	prLevelmonthMA12	0.391966792	0.00000000000000000652158941	0.252 < prLevelmonthMA12 <= 0.427	list(vol_5MA = 0.349135571180205, tradeVolume = 0.345...	list(Yes = 0.612919688224792, No = 0.38708034157753)

- model_type：模型類別，此處為分類
- case：分析樣本編號
- label：分類標籤，此處Yes為漲，No為跌
- label_prob：預測分類的機率，即為Keras模型預測漲的機率
- model_r2：線性模型的R^2，即模型的解釋能力
- model_intercept：模型截距項
- model_prediction：模型預測結果，即為Keras模型預測漲的機率

- feature：重要特徵名稱
- feature_value：特徵值
- feature_desc：特徵選取範圍
- data：原始資料
- prediction：預測分類的機率，同Keras模型

利用關聯分析來解讀LIME模型結果

- 由於LIME模型是針對每筆挑選的樣本進行分析，所以每筆樣本選出來的重要特徵皆會不一樣
- 要分析LIME模型的結果，可利用關聯分析法則找出常見的特徵組合

case	feature_desc
1	0.332 < vol_5MA <= 0.388
1	0.740 < weekMACD
1	0.569 < dayBBdn
1	0.333 < bottom4PricePer <= 0.500
1	prLeveldayMA90 <= 0.258
2	0.332 < vol_5MA <= 0.388
2	dayFastD <= 0.227
2	0.382 < dayBBup <= 0.571
2	0.290 < monthBBup <= 0.384
2	0.252 < prLevelmonthMA12 <= 0.427
3	0.388 < vol_5MA <= 0.518
3	0.532 < slopeBottom46 <= 0.606
3	0.262 < prLeveldayMA60 <= 0.479
3	prLeveldayMA20 <= 0.396
3	0.258 < prLeveldayMA90 <= 0.469
4	0.388 < vol_5MA <= 0.518
4	0.577 < monthBBup
4	0.285 < weekMACD <= 0.507
4	0.160 < monthFastD <= 0.541
4	daySlowD <= 0.230
5	0.332 < vol_5MA <= 0.388
5	0.763 < dayFastD
5	bottom4DatePer <= 0.219
5	0.541 < monthFastD <= 0.838
5	0.675 < prLevelweekMA20

利用關聯分析來解讀LIME模型結果

- 由於每年的模型皆不一樣，所以需要對各年的模型進行分析，此處以2017年作範例

```r
# 利用購物籃分析
library(arules)
data <- limeAnalysisData %>%
  filter(predictYear=="2017") %>%     # 選擇2017年的LIME模型結果進行分析
  mutate(case=as.numeric(case)) %>%
  select(case, feature_desc)
data <- as(split(data$feature_desc, data$case),"transactions")        # 轉換為關聯法則資料格式
rules <- apriori(data, parameter=list(support=0.03,confidence=0.05))  # 執行關聯法則
output <- inspect(rules)                                               # 轉換為data.frame格式
colnames(output)[2] <- "assign"                                        # 更換欄位名稱
output <- output %>% arrange(desc(confidence))                         # 以信心水準由大到小排序
```

利用關聯分析來解讀LIME模型結果

- 2017年分析結果

lhs	assign	rhs	support	confidence	lift	count
{0.500 < weekSignal <= 0.745}	=>	{0.332 < vol_5MA <= 0.388}	0.04054054	1.00000000	2.387097	3
{slopeBottom26 <= 0.417}	=>	{0.332 < vol_5MA <= 0.388}	0.04054054	1.00000000	2.387097	3
{prLeveldayMA5 <= 0.290}	=>	{0.332 < vol_5MA <= 0.388}	0.04054054	1.00000000	2.387097	3
{0.333 < bottom4PricePer <= 0.500}	=>	{0.332 < vol_5MA <= 0.388}	0.05405405	1.00000000	2.387097	4
{0.577 < monthBBup}	=>	{0.388 < vol_5MA <= 0.518}	0.08108108	0.85714286	2.883117	6
{0.165 < monthFastK <= 0.540}	=>	{0.332 < vol_5MA <= 0.388}	0.04054054	0.75000000	1.790323	3
{0.569 < dayBBmavg}	=>	{0.332 < vol_5MA <= 0.388}	0.04054054	0.60000000	1.432258	3
{prLeveldayMA10 <= 0.290}	=>	{0.332 < vol_5MA <= 0.388}	0.04054054	0.60000000	1.432258	3
{0.430 < prLevelweekMA26 <= 0.649}	=>	{0.332 < vol_5MA <= 0.388}	0.04054054	0.60000000	1.432258	3
{}	=>	{0.332 < vol_5MA <= 0.388}	0.41891892	0.41891892	1.000000	31
{}	=>	{0.388 < vol_5MA <= 0.518}	0.29729730	0.29729730	1.000000	22
{0.388 < vol_5MA <= 0.518}	=>	{0.577 < monthBBup}	0.08108108	0.27272727	2.883117	6
{}	=>	{0.518 < vol_5MA}	0.16216216	0.16216216	1.000000	12
{0.332 < vol_5MA <= 0.388}	=>	{0.333 < bottom4PricePer <= 0.500}	0.05405405	0.12903226	2.387097	4
{}	=>	{vol_5MA <= 0.332}	0.12162162	0.12162162	1.000000	9
{0.332 < vol_5MA <= 0.388}	=>	{0.500 < weekSignal <= 0.745}	0.04054054	0.09677419	2.387097	3
{0.332 < vol_5MA <= 0.388}	=>	{slopeBottom26 <= 0.417}	0.04054054	0.09677419	2.387097	3
{0.332 < vol_5MA <= 0.388}	=>	{prLeveldayMA5 <= 0.290}	0.04054054	0.09677419	2.387097	3
{0.332 < vol_5MA <= 0.388}	=>	{0.165 < monthFastK <= 0.540}	0.04054054	0.09677419	1.790323	3
{0.332 < vol_5MA <= 0.388}	=>	{0.569 < dayBBmavg}	0.04054054	0.09677419	1.432258	3
{0.332 < vol_5MA <= 0.388}	=>	{prLeveldayMA10 <= 0.290}	0.04054054	0.09677419	1.432258	3
{0.332 < vol_5MA <= 0.388}	=>	{0.430 < prLevelweekMA26 <= 0.649}	0.04054054	0.09677419	1.432258	3