

$$(X, O) = e^{-\frac{x^2}{2\sigma^2}}$$

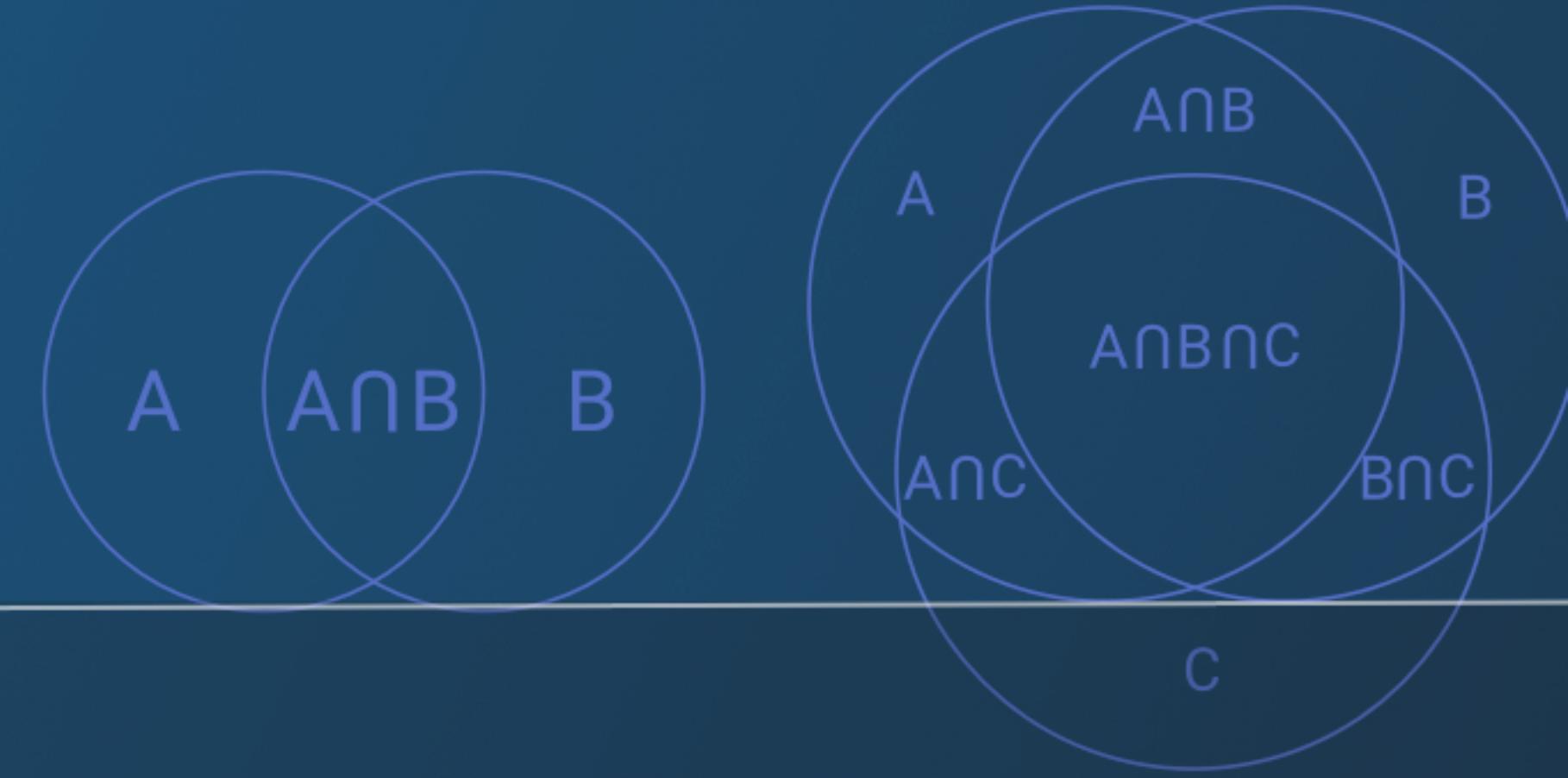
$$x(X, O) = -\frac{x}{\sigma^2} G(X, O) = -\frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$$

$$xx(X, O) = \frac{x^2 - \sigma^2}{\sigma^4} G(X, O) = \frac{x^2 - \sigma^2}{\sigma^4} e^{-\frac{x^2}{2\sigma^2}}$$

$$xxx(X, O) = -\frac{x^3 - x\sigma^2}{\sigma^6} G(X, O) = -\frac{x^3 - x\sigma^2}{\sigma^6} e^{-\frac{x^2}{2\sigma^2}}$$

Julia 程式語言學習馬拉松

Day 18



$$\ln(x + \sqrt{1+x^2}) + x - \frac{1}{x + \sqrt{1+x^2}} \left(1 + \frac{x}{\sqrt{1+x^2}} \right)$$



cupay 陪跑專家 : James Huang

DataFrames.jl 介紹 (二): Joins 與 Split-Apply- Combine Strategy





重要知識點



- 在關聯式資料庫的操作中，查詢時常使用到 Join，例如 Inner 、 Left 、 Right 、 Outer 、 Semi 、 Anti 、 Cross Join ，在 DataFrame 中均提供 Join 的操作。
- Split-Apply-Combine 策略框架 (大陸有人譯做“化整為零策略”) 指的是針對資料進行 group by 、 map 、 combine ，依行 (column) 或列 (row) 進行資料處理；或將資料分組之後，進行獨立的操作；然後將處理完成的片段進行組合。
- 在範例程式中會對於 DataFrames.jl 如何實作及應用進行示範。



DataFrame : Join



Join	語法範例	說明
Inner	<code>join(df1, df2, on = :ID, kind = :inner)</code>	結果包含符合條件的資料及所有欄位.
Left	<code>join(df1, df2, on = :ID, kind = :left)</code>	結果包含 df1 所有資料, 但若未對應到的 df2 欄位值則填入 missing.
Right	<code>join(df1, df2, on = :ID, kind = :right)</code>	結果包含 df2 所有資料, 但若未對應到的 df1 欄位值則填入 missing.
Outer	<code>join(df1, df2, on = :ID, kind = :outer)</code>	結果包含 df1 和 df2 所有資料, 未對應到的欄位值則填入 missing.
Semi	<code>join(df1, df2, on = :ID, kind = :semi)</code>	跟 Inner Join 相同, 但是結果僅包含 df1 的欄位.
Anti	<code>join(df1, df2, on = :ID, kind = :anti)</code>	結果包含僅存在於 df1 的資料, 且僅包含 df1 的欄位.
Cross	<code>join(df1, df2, kind = :cross, makeunique = true)</code>	為 df1 X df2 的結果, 為 2 個 DataFrame 的乘積.



DataFrame : Join



- 若是要 Join 連接的 Key 值欄位名稱不一樣的話，可以在 on 參數指明其對應，例如：`join(df1, df2, on = :ID => :IDnew)`。
- indicator 參數可以在 Join 結果中加入顯示資料是存在於 left 或是 right 或是 both (兩個 DataFrame 均存在)。
- 若要檢核要連接的 Key 值是否有重複，可以使用 validate 參數檢核 left 或 right (或是兩邊都檢核)，檢核指定時可以使用 tuple 或是 pair，例如：`validate = (true, true)`。



The Split-Apply-Combine Strategy



- R 及資料科學大神 Hadley Wickham 在 2011 年出版了 The Split-Apply-Combine Strategy for Data Analysis，提出資料處理的策略框架。
- 說明針對資料進行 group by、map / apply (R 語言)、combine，依行 (column) 或列 (row) 進行資料處理；或將資料分組之後，進行獨立的操作；然後將處理完成的片段進行組合。
- DataFrames.jl 支援這個框架，以 by() 函式來實作 Split-Apply-Combine 策略。

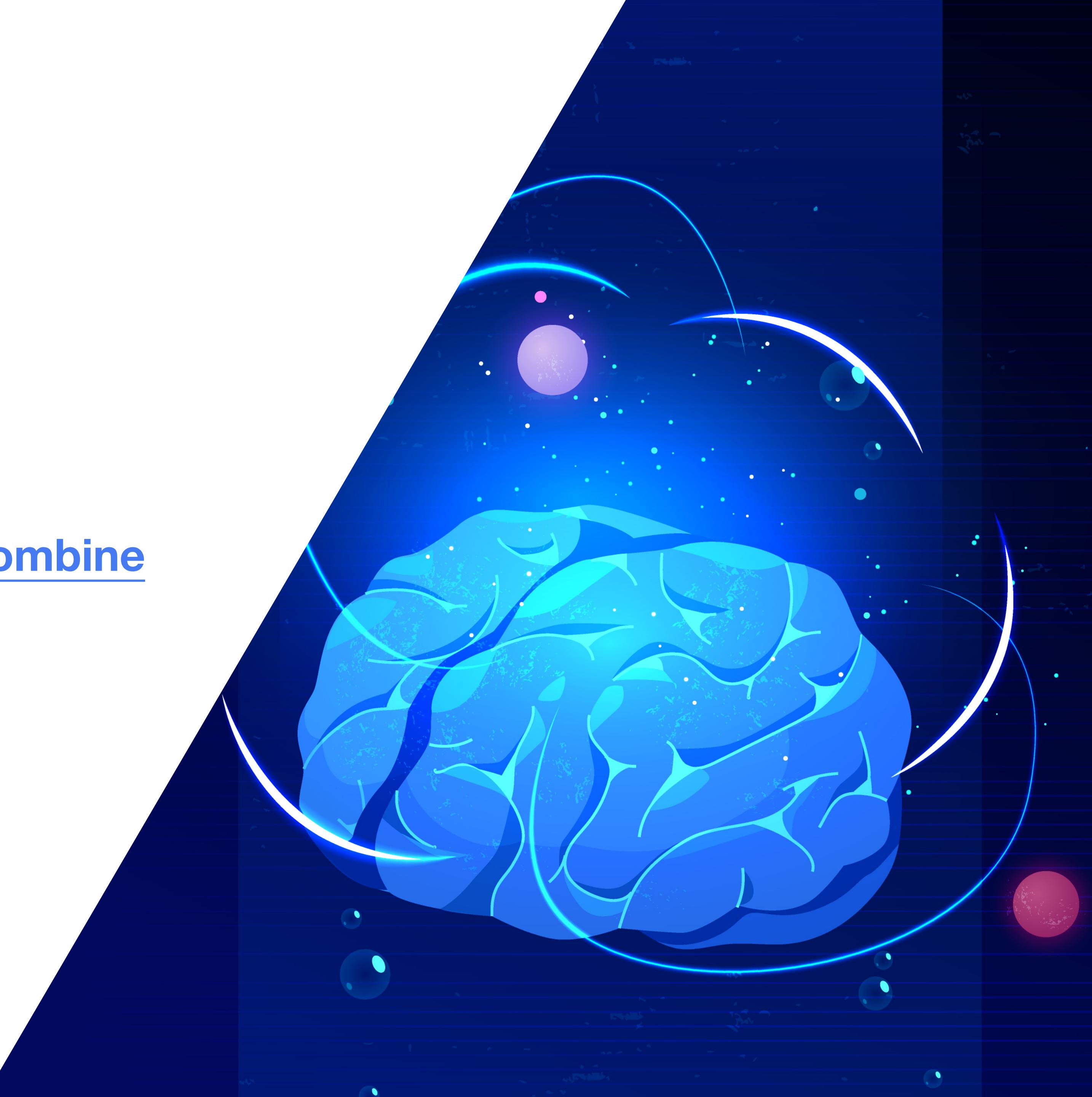
知識黑點 回顧

- 延伸關聯式資料庫的 Join，DataFrames.jl 也提供相對應的 Join 函式功能。
- 透過 Join 可以連接合併 2 個 DataFrame；使用不同的 Join 可能產生不同的結果。
- Split-Apply-Combine 讓資料處理的操作上更有效率也更方便，在不同的程式語言上語法略有不同，但在 Julia / DataFrame.jl 中則是以 by() 函式來實作。



推薦閱讀

- [Wikipedia 連接 \(Join\) 的介紹](#)
- [Hadley Wickham “The Split-Apply-Combine Strategy for Data Analysis”](#)





解題時間

請跳出 PDF 至官網 Sample Code
& 作業開始解題