

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

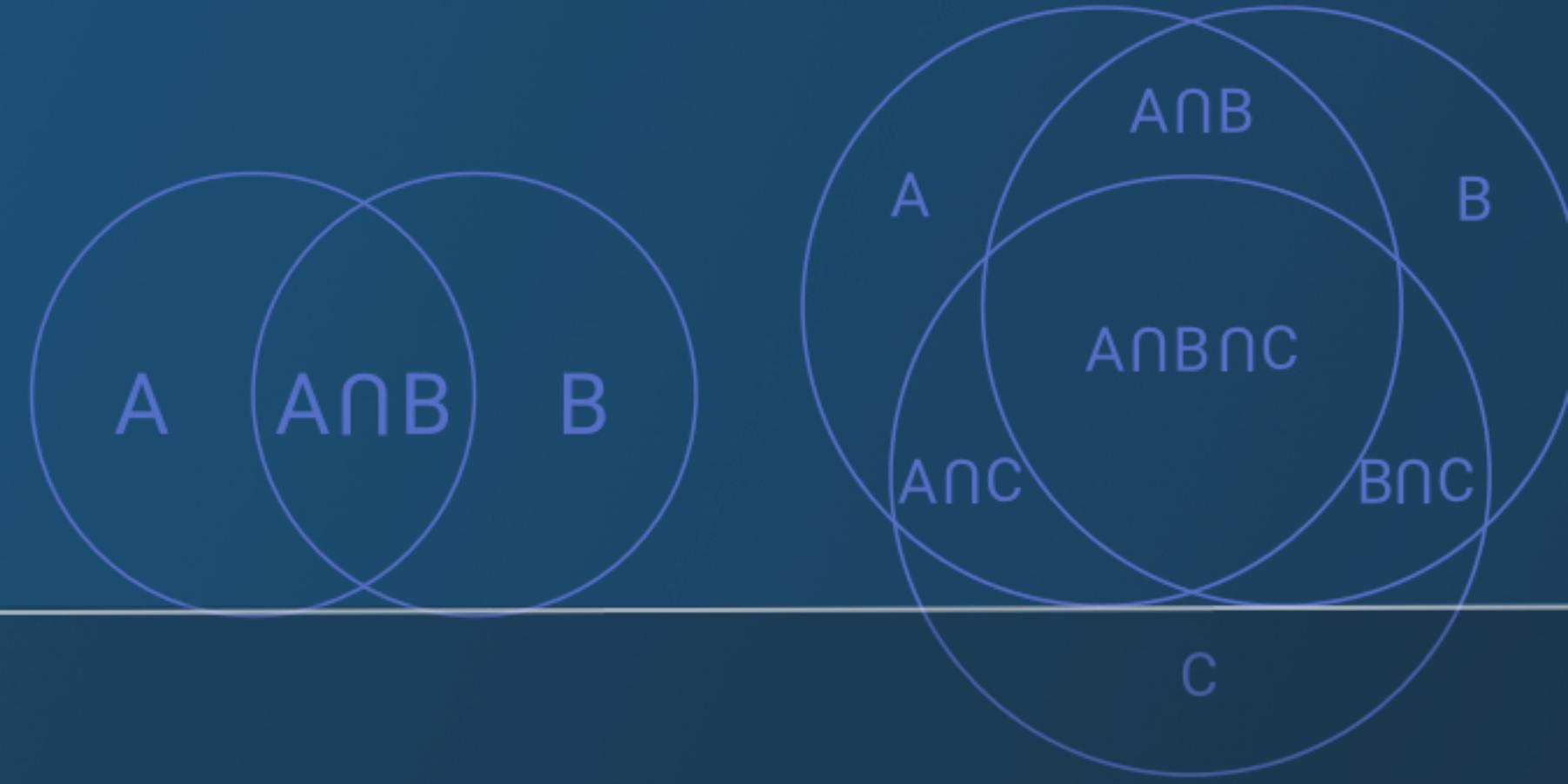
$$x(X, O) = -\frac{x}{\sigma^2} G(X, O) = -\frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$$

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

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

Julia 程式語言學習馬拉松

Day 27



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



cupay 陪跑專家 : Andy Tu

GLM 線性迴歸





重要知識點



- 了解 GLM 套件所提功能。
- 了解單純線性迴歸及多重線性迴歸模型。
- 了解廣義線性模型。



GLM 套件



- GLM 套件提供了廣義線性模型的各項功能，當中包含線性迴歸模型。
- 廣義線性模型會在後續介紹。
- 請先自行安裝 GLM 套件。

```
]add GLM
```



準備資料集

- 本教材的示範資料集皆來自 RDatasets。

```
using RDatasets  
data = RDatasets.dataset("datasets",  
"longley")
```



單純線性迴歸



- 單純線性迴歸 (simple linear regression) 可以說是最簡單的機器學習模型了。
- 它假設了輸入與輸出之間為線性關係，而輸入的變數只有一個。
- `@formula(Employed ~ GNP)` 可以讓使用者撰寫線性方程式，在 ~ 左邊的是要預測的變項，在右邊的是輸入的變項，直接以 DataFrame 的欄位名稱為變數名稱即可。

```
model = GLM.lm(@formula(Employed ~  
GNP), data)
```



模型訓練



- GLM.lm 為訓練及建構線性迴歸模型的函式，第一個參數給定方程式，第二個參數給定資料。
- 函式執行完後就等同於訓練結束了。

```
model = GLM.lm(@formula(Employed ~  
GNP), data)
```



模型預測

- 訓練完成的模型可以使用 predict 來做預測，第一個參數給定訓練過的模型，第二個參數給定新的資料。

```
new_X = DataFrame(GNP=[300., 400.,
500.])
predict(model, new_X)
```



模型評估

- 一般評估線性迴歸模型的好壞會使用 R^2 值，數值範圍從 0~1。
- 越接近 1 是越好，代表資料越靠近線性模型。
- 也有校正的 adjusted R^2 值可以使用。

```
r2(model)  
adjr2(model)
```



模型參數

- 要取得模型參數可以使用 `coef`。

```
intercept, slope = coef(model)
```



多重線性迴歸

- 多重線性迴歸 (multiple linear regression) 我們採用另外一個資料集。
- 模型的建構與訓練跟前面的相似，只是將輸入變項增多。

```
data = RDatasets.dataset("datasets",
"mtcars")
model = GLM.lm(@formula(MPG ~ Cyl +
Disp + HP + DRat + WT + QSec + VS +
AM + Gear + Carb), data)
```



其他迴歸模型

- 在 GLM 所支援的迴歸模型中還有很多：
 - 羅吉斯迴歸 (Logistic regression)
 - 卜瓦松迴歸 (Poisson regression)
 - 負二項式迴歸 (Negative binomial regression)
 - ...



廣義線性模型

- 多樣的迴歸模型可以入統一的廣義線性模型（Generalized linear model）框架中。
- 廣義線性模型需要使用 `glm` 函式，可以利用 `glm` 加上不同的分佈函數及鏈結函數組合出不同的迴歸模型。
- 第一、二個參數的使用與 `lm` 一樣，`dist` 為分佈函數，`link` 為鏈結函數。

```
model = glm(@formula(), data, dist, link)
```



鏈結函數

- CauchitLink
- CloglogLink
- IdentityLink
- InverseLink
- InverseSquareLink
- LogitLink
- LogLink
- NegativeBinomialLink
- ProbitLink
- SqrtLinkr



典型分佈模型

- Bernoulli : 一般搭配 LogitLink
- Binomial : 一般搭配 LogitLink
- Gamma : 一般搭配 InverseLink
- InverseGaussian : 一般搭配 InverseSquareLink
- NegativeBinomial : 一般搭配 LogLink
- Normal : 一般搭配 IdentityLink
- Poisson : 一般搭配 LogLink

知識點 回顧

- 了解 GLM 套件。
- 學會使用（單純）線性迴歸模型。
- 了解廣義線性模型的使用及其他不同的迴歸模型。



推薦閱讀

- [GLM.jl 官方文件](#)





解題時間

請跳出 PDF 至官網 Sample Code
& 作業開始解題