自我評量與練習

第1章 認識資料科學

1. (2) 下列哪一種情況屬於對於資料的使用陷入誤用預測的狀況？(1) 特定資料的蒐集與交換讓個人的隱私暴露 (2) 以數據分析的結果來判定一個人的罪責或犯罪動機(3) 以年收入來當做政府補助低收入戶的門檻，未能真正照顧有需求者 (4) 以過去的消費行為來推敲消費者的喜好。

解析：請參考1-1-2節的內容

1. (3) 下列哪一項特性不在大數據6V的特性中？(1) 價值(value) (2) 準確性(veracity) (3) 揮發性(volatility) (4) 改變性(variability)。

解析：請參考1-2-1節的內容

1. (4) 「在工作能完成的時限內，工作量會一直增加，直到所有的可用時間都被填充為止，可引申說明儲存裝置上的空間會一直被資料填滿」，以上描述的是下列那一項？(1) 辛普森悖論(Simpson’s paradox) (2) 安斯庫姆四重奏(Anscombe’s Quartet) (3) 黑天鵝效應(Black swan theory) (4) 帕金森定理(Parkinson’s law)。

解析：請參考1-2-2節的內容

1. (3) 下列哪一種等級的資料量最大？(1) TB (2) PB (3) EB (4) GB。

解析：請參考1-2-4節的內容

1. (2) 影像、視訊屬於下列哪一類的資料？(1) 結構化的資料(structured data) (2) 非結構化的資料(unstructured data) (3) 半結構化的資料(semi-structured data) (4) 表面結構化的資料(quasi-structured data)。

解析：請參考1-2-4節的內容

第2章 數據分析的應用

1. (1) 下列哪一項描述未必一定正確？(1) 資料量越大能發揮的效果也越大 (2) 資料可能要經過一段時間才能完整地產生 (3) 大數據分析所需要的資料未必一定是完整的 (4) 小數據的分析還是可能有其特定的用途。

解析：請參考2-1-1節的內容

1. (3) 下列有關於資料運用的描述，哪一項是錯誤的？(1) 相關性重於因果關係(2) 掌握龐大的整體資料是關鍵 (3) 要找出因果關係才能發揮作用 (4) 真實的資料是雜亂的。

解析：請參考2-1-2節的內容

1. (1) 下列有關於資料科學成功運用關鍵的描述，哪一項是錯誤的？(1) 決策思維的轉變應以個人經驗與直覺為主，資料分析為輔 (2) 資料策略上應充分了解所解決的問題 (3) 要注重數據分析與管理的專業養成 (4) 應正確導入與解讀數據。

解析：請參考2-2-4節的內容

1. (3) 下列有關於資料使用風險的描述，哪一項是錯誤的？(1) 資料的蒐集、處理與運用，應該符合個資保護的要求 (2) 不管是有意或是無意的誤用都有可能產生傷害 (3) 資料經過遮罩處理以後就不會有個資外洩的問題 (4) 假如資料不是真實的，不僅無法提供用途，還可能產生負面的影響。

解析：請參考2-4節的內容，不恰當的遮罩處理以後還是可能因為經由資料的推理比對找出個資。

1. (1) 下列有關於資料科學應用的描述，哪一項是錯誤的？(1) 數據分析預測的結果會與一般多數人的直覺相近 (2) 預測未來常會用到過去的資料 (3) 資料量的多寡跟資料的價值不一定相關 (4) 數據分析的預測不一定都需要完整的資料。

解析：請參考2-3-7節的內容

第3章 相關的資訊技術

1. (3) 「讓使用者在自己的電腦上連接雲端伺服器，執行伺服器上的應用軟體」，以上描述的是下列哪一種雲端服務的模式？(1) IaaS (2) PaaS (3) SaaS (4) DaaS。

解析：請參考3-2-3節的內容

1. (1) 「由特定組織擁有，雲端基礎設施提供雲端服務給一般大眾或產業」，以上描述的是下列哪一種雲端運算的部署模型？(1) public cloud (2) private cloud (3) community cloud (4) hybrid cloud。

解析：請參考3-2-3節的內容

1. (4) 下列關於虛擬化技術的描述，哪一項是錯誤的？(1) 虛擬機器(virtual machine)和母系統(host system)或是其他的虛擬機器是分隔的 (2) 虛擬機器可以被打包成單一的檔案 (3) 所有的硬體資源可當成是一個資源池 (4) 無法彈性支援效能高的資訊環境。

解析：請參考3-1-1節的內容

1. (3) 「也稱為軟體定義網路(software-defined networking)，主要是把實體網路當成連線的資源池」，以上描述的是下列哪一種虛擬化技術？(1) 伺服器虛擬化(server virtualization) (2) 應用虛擬化(application virtualization) (3) 網路虛擬化(network virtualization) (4) 處理器虛擬化(processor virtualization)。

解析：請參考3-2-2節的內容

1. (1) 「在只有一台實體伺服器的情況下，支援多台虛擬伺服器的運作」，以上描述的是下列哪一種虛擬化技術？(1) 伺服器虛擬化(server virtualization) (2) 應用虛擬化(application virtualization) (3) 網路虛擬化(network virtualization) (4) 處理器虛擬化(processor virtualization)。

解析：請參考3-2-2節的內容

第4章 常用的語言與工具

1. (4) 假如在程式中定義的函數使用的參數採用傳值呼叫(call by value)的方式傳遞，則函數中對於參數的處理跟主程式中對應的變數會有什麼影響？(1) 會改變主程式中對應的變數的值 (2) 會改變主程式中對應的變數的位址 (3) 會同時改變主程式中對應的變數的值與位址 (4) 不會產生影響。

解析：請參考4-4-1節的內容

1. (2) 下列關於程式語言中遞迴(recursion)的描述，哪一項是錯誤的？(1) 遞迴的語法可以簡化程式 (2) 遞迴的語法在程式設計上建議常用 (3) 遞迴是指函數呼叫自己 (4) 遞迴的執行相當耗費運算資源。

解析：請參考4-4-1節的內容，遞迴的語法在程式設計上並不建議常用

1. (2) 像TSP(Traveling Salesperson)這樣的問題，在微軟的Excel軟體中可以使用什麼功能來得到答案？(1) 加總計算 (2) 規劃求解 (3) 統計函數 (4) 字串處理函數。

解析：請參考4-1-2節的內容

1. (3) 下列關於Python語言的描述，哪一項是錯誤的？(1) 在資料科學與大數據的領域很受歡迎 (2) Anaconda提供了好幾個與Python相關的工具 (3) 大多數人學Python的目的在於開發複雜的資訊系統 (4) Python在資料科學與大數據的領域有很多相關的套件。

解析：請參考4-2-1節的內容

1. (3) 下列哪一項通常不是指可再使用程式的集合？(1) 套件(package) (2) 函數庫(library) (3) 主程式(main program) (4) 模組(module)。

解析：請參考4-4-2節的內容

第5章 進行數據分析

1. (3) 下列關於統計觀念的描述，哪一項是錯誤的？(1) 樣本(sample)是母體中的部分成員，相當於數學中所說的子集合(subset) (2) 母體的成員未必一定是人(3) 針對樣本得到的統計結果，可以完全適用於母體(4) 母體(population)是指我們要探討的事物的全體。

解析：請參考5-2-1節的內容

1. (4) 下列關於機率(probability)的描述，哪一項是錯誤的？(1) 執行特定的動作並觀察結果，這在機率理論中稱為試驗(trial) (2) 進行試驗時，符合特定條件的結果形成的集合稱為事件(event) (3) 機率可定義為在一個動作的樣本空間中，發生特定事件的可能性 (4) 丟硬幣會出現正面的機率是40%。

解析：請參考5-2-2節的內容

1. (4) 在進行統計的檢定時常假設資料的分佈是下列哪一種分佈？(1) 雜亂分佈 (2) 二元分佈 (3) 隨機分佈 (4) 常態分佈。

解析：請參考5-1-1節的內容

1. (1) 下列關於異常值(outlier)的描述，哪一項是錯誤的？(1) 對於整體的資料來說，異常值對於平均值(mean)不會有影響 (2) 對於整體的資料來說，異常值對於中位數(median)沒有影響 (3) 以統計的定義來看，異常值(outlier)是以高於或低於平均數兩個標準差(standard deviation)來判斷 (4) 對於某些應用來說，異常值可能反映一些特殊事件的發生。

解析：請參考5-1-2節的內容

1. (4) 探討如何促使特定的現象發生在數據分析領域是屬於哪一種數據分析的方法？(1) descriptive analytics (2) diagnostic analytics (3) predictive analytics (4) prescriptive analytics。

解析：請參考5-3-1節的內容

第6章 機器學習與人工智慧

1. (1) 「使用「移動中的資料(data in motion)」，採用的模型會隨著收到的資料而改變」，以上描述的是下列哪一種機器學習的方法？(1) 線上學習 (2) 轉移式的學習(3) 整體性的學習 (4) 深度學習。

解析：請參考6-3節的內容

1. (2) 「沒預期要驗證的目標，通常要嘗試從資料中找出規律(pattern)，或是發現隱藏的事實」，以上描述的是下列哪一種機器學習的模型？(1) supervised learning (2) unsupervised learning (3) semi-supervised learning (4) reinforcement learning。

解析：請參考6-1-2節的內容

1. (3) 下列哪一項比較不算是機器學習中會進行的主要工作？(1) 資料探索分析 (2) 特徵工程 (3) 軟體工程的程序 (4) 資料清理。

解析：請參考6-1-1節的內容

1. (3) 下列關於機器學習中特徵工程的描述，哪一項是錯誤的？(1) 數據分析的成敗，跟特徵的萃取有密切的關係 (2) 有時某些特徵沒有直接存在資料裡，譬如客戶的忠誠度 (3) 透過資料試著找出特徵的過程就是特徵工程(feature engineering)，這是機器學習裡頭比較單純簡易的工作 (4) 所謂的「特徵(feature)」可能只是一個數字或是類別，從資料萃取出來，描述特定的屬性。

解析：請參考6-1-3節的內容

1. (3) 「使用多種演算法，整合多個較弱的學習模型有可能產生一個較強而準備的學習模型」，以上描述的是下列哪一種機器學習的方法？(1) 線上學習 (2) 轉移式的學習(3) 整體性的學習 (4) 深度學習。

解析：請參考6-3節的內容

第7章 延伸學習與職涯規劃

1. (2) 下列哪一個MOOCS平台是由國內的機構負責營運的？(1) edx (2) ewant (3) Coursera (4) Udemy。

解析：請參考7-1-1節的內容

1. (4) 下列哪一個課程比較不屬於與資料科學直接相關的課程？(1) 線性代數 (2) 資料採礦 (3) 機率 (4) 策略管理。

解析：請參考7-1-2節的內容

1. (4) 下列那一項工作職稱所指的的是機構內資料科學管理與運用的最高主管？(1) 執行長(chief executive officer) (2) 財務長(chief financial officer) (3) 數據長(chief data officer) (4) 資安長(chief security officer)。

解析：請參考7-2-3節的內容

1. (2) 「管理公司的資料庫系統，綜理資料的日常運作，監督資料的備份與稽核」，以上的工作內容描述與下列哪一種職缺的要求最接近？(1) 資料分析師 (2) 資料管理師 (3) 系統工程師 (4) 系統分析師。

解析：請參考7-2-3節的內容

1. (3) 下列哪一種方式對於資料科學思維的建立可能比較沒有幫助？(1) 多了解各種資料科學的應用 (2) 參與資料科學的專案工作 (3) 撰寫Python程式來處理字串資料 (4) 與有資料分析需求的客戶進行溝通互動。

解析：請參考1-2-4節的內容