<table>
<tr><th>위치</th><th>오류유형</th><th>수정 전</th><th>수정 후</th></tr>
<tr><td>46p
번호 : 11</td><td>문제-문항</td><td>① 제품과 서비스의 유형</td><td>① 거래의 형태</td></tr>
<tr><td>61p</td><td>문제-본문</td><td>
<table>
<tr><th colspan="2">유 형</th><th>데이터 소스</th></tr>
<tr><td colspan="2">정형 데이터</td><td>• 업무 처리, 매개거래, 로그 데이터, 시계열 데이터
• DB to DB, Sqoop, EAI(Enterprise Application Intergration)
• ETL(Extract, Transform, Load) 등을 이용한 수집</td></tr>
<tr><td colspan="2">반정형 데이터</td><td>• 데이터 내부에 데이터 구조에 대한 메타 정보 포함
• HTML, XML, JSON, RSS, 웹로그, 센서 데이터 등</td></tr>
<tr><td rowspan="4">비정형 데이터</td><td>일 반</td><td>• 소셜 미디어(SNS), 고객 서비스, 품질보증
• 이벤트, 이메일, 자유형식의 텍스트</td></tr>
<tr><td>센서 데이터</td><td>온도, QR 코드, RFID, GPS</td></tr>
<tr><td>새로운 데이터 유형</td><td>GPS 매칭, 오디오, 정지화상, 비디오</td></tr>
<tr><td>데이터 수집</td><td>Crawler, FTP, HTTP Protocol 수집, Parsing</td></tr>
</table>
</td><td>
<table>
<tr><th colspan="2">유 형</th><th>데이터 소스</th></tr>
<tr><td colspan="2">정형 데이터</td><td>• 업무 처리, 매개거래, 로그 데이터, 시계열 데이터
• DB to DB, Sqoop, EAI(Enterprise Application Intergration)
• ETL(Extract, Transform, Load) 등을 이용한 수집</td></tr>
<tr><td colspan="2">반정형 데이터</td><td>• 데이터 내부에 데이터 구조에 대한 메타 정보 포함
• HTML, XML, JSON, RSS, 웹로그, 보안 및 특정 센서 데이터 등</td></tr>
<tr><td rowspan="4">비정형 데이터</td><td>일 반</td><td>• 소셜 미디어(SNS), 고객 서비스, 품질보증
• 이벤트, 이메일, 자유형식의 텍스트</td></tr>
<tr><td>센서 데이터</td><td>온도, QR 코드, RFID, GPS</td></tr>
<tr><td>새로운 데이터 유형</td><td>GPS 매칭, 오디오, 정지화상, 비디오</td></tr>
<tr><td>데이터 수집</td><td>Crawler, FTP, HTTP Protocol 수집, Parsing</td></tr>
</table>
</td></tr>
<tr><td>66p</td><td>문제-본문</td><td>
<table>
<tr><th>유 형</th><th>데이터 종류</th><th>저장 시스템</th></tr>
<tr><td>정 형</td><td>RDB, 스프레드시트</td><td>RDB</td></tr>
<tr><td>반정형</td><td>HTML, XML, JSON, 웹문서, 웹로그, 센서 데이터</td><td>RDB, NoSQL</td></tr>
<tr><td>비정형</td><td>소셜 데이터, 문서(워드, 한글), 이미지, 오디오, 비디오, IoT</td><td>NoSQL, 분산파일시스템(HDFS)</td></tr>
</table>
</td><td>
<table>
<tr><th>유 형</th><th>데이터 종류</th><th>저장 시스템</th></tr>
<tr><td>정 형</td><td>RDB, 스프레드시트</td><td>RDB</td></tr>
<tr><td>반정형</td><td>HTML, XML, JSON, 웹문서, 웹로그, 보안 및 특정 센서 데이터</td><td>RDB, NoSQL</td></tr>
<tr><td>비정형</td><td>소셜 데이터, 문서(워드, 한글), 이미지, 오디오, 비디오, IoT</td><td>NoSQL, 분산파일시스템(HDFS)</td></tr>
</table>
</td></tr>
<tr><td>67p</td><td>문제-본문</td><td>Google GFS(Google File System)</td><td>Google의 GFS(Google File System)</td></tr>
<tr><td>152p
번호 : 53</td><td>해설</td><td>정답 ④
해설 키, 온도는 등간 척도, 순위는 서열 척도이다.</td><td>정답 ②, ④
해설 온도는 등간 척도, 순위는 서열 척도이다.</td></tr>
<tr><td>204p
번호 : 38</td><td>문제-문항</td><td>④ MapReduce</td><td>④ MapR</td></tr>
<tr><td>226p</td><td>문제-본문</td><td>상자그림 그리는 방법
• 위의 [그림]은 11명의 중간고사 성적(89점, 88점, 67점, 78점, 90점, 88점, 78점, 90점, 86점, 55점)에 대해서 상자그림을 그린 것이다. 상자그림을 그리는 방법은 다음과 같다.</td><td>상자그림 그리는 방법
• 위의 [그림]은 11명의 중간고사 성적(89점, 88점, 67점, 78점, 90점, 88점, 78점, 90점, 86점, 55점, 78점)에 대해서 상자그림을 그린 것이다. 상자그림을 그리는 방법은 다음과 같다.</td></tr>
<tr><td>230p</td><td>문제-본문</td><td>$s = \sqrt{\frac{1}{(n-1)}\left(\sum x_i^2 - n\overline{x^2}\right)}$</td><td>$s = \sqrt{\frac{1}{(n-1)}\left(\sum x_i^2 - n\overline{x}^2\right)}$</td></tr>
</table>

위치	오류유형	수정 전	수정 후
266p	문제-본문	㉡ 공차한계가 0.1 이하일 때 다중공선성 문제가 **없다고** 판정한다(분산팽창지수의 역수값).	㉡ 공차한계가 0.1 이하일 때 다중공선성 문제가 **있다고** 판정한다(분산팽창지수의 역수값).
311p	문제-본문	① $\overline{x} = 77.5$(표본평균)	① $\overline{x} = 76.7$(표본평균)
313p	문제-본문	**⑤ 신뢰구간 예시** 상기 통학시간 자료의 95% 신뢰구간을 구하시오.	**⑤ 신뢰구간 예시** 다음 통학시간(분) 자료의 95% 신뢰구간을 구하시오. (60, 65, 50, 45, 35, 40, 45, 60, 50, 50)
337p	문제-본문	표본 ㅣ	표본 1
410p	문제-본문	$= r\frac{s_y}{s_x}$	$\hat{\beta} = r\frac{s_y}{s_x}$
458p	문제-본문	$\beta_0 = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sum(x_i - \overline{x})}$	$\beta_1 = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sum(x_i - \overline{x})^2}$

위치	오류유형	수정 전	수정 후
474~475p	문제-본문	**장 점** ㉠ 첫째, 다양한 형태의 데이터에 적용이 가능하고 분석방법의 적용이 용이하다. ㉡ 둘째, 이해하고 구현하기가 쉬우며 결과 해석이 용이하다는 것이다. **단 점** ㉠ 첫째, 초기 군집수를 결정하는 것이 어렵고 결과 해석이 어렵다는 것이다. ㉡ 둘째, 데이터의 평균값 정의가 가능한 데이터에 대해서만 사용이 가능하다는 것이다. ㉢ 셋째, 아웃라이어에 매우 민감하다는 것이다.	**장 점** ㉠ 알고리즘이 쉬워 구현이 비교적 쉽다. ㉡ 주어진 알고리즘으로 대규모 적용이 가능하다. ㉢ 다양한 형태의 데이터에 적용이 가능하다. **단 점** ㉠ 초기 k를 찾기 어렵다. ㉡ 모든 점에서 계산하기 때문에 속도가 느리다. ㉢ 평균을 사용하기 때문에 아웃라이어에 영향을 많이 받는다. ㉣ 가중치와 거리정의가 어렵다. ㉤ 사전에 주어진 목적이 없으므로 결과를 해석하는 데 있어 어려움이 존재한다.
478p	문제-본문	$S_{yy} = \sum (y_i - y)^2$	$S_{yy} = \sum (y_i - \overline{y})^2$
478p	문제-본문	$S_{xy} = \sum (x_i - \overline{x})(y_i - y)$	$S_{xy} = \sum (x_i - \overline{x})(y_i - \overline{y})$
590p	문제-본문	$\text{중앙값} = \begin{cases} x_{(n+1)/2}, & n\text{이 홀수} \\ x_{n/2}, & n\text{이 짝수} \end{cases}$	$\text{중앙값} = \begin{cases} x_{(n+1)/2}, & n\text{이 홀수} \\ \frac{1}{2}(x_{n/2} + x_{(n/2+1)}), & n\text{이 짝수} \end{cases}$
608p	문제-본문	• 지지도(Support) : 전체 데이터세트에서 해당 아이템 집합이 포함된 비율 $S(X) = \frac{Count(X)}{N} =, S(X, Y) = \frac{Count(X, Y)}{N} = P(X \cap Y)$	지지도(Support) : 전체 데이터세트에서 해당 아이템 집합이 포함된 비율 $S(X) = \frac{Count(X)}{N} = P(X), S(X, Y) = \frac{Count(X, Y)}{N} = P(X \cap Y)$ 즉, 지지도 $S(X)$는 N명의 모집단에서 사건 X가 일어날 확률의 개념과 유사하게 정의됨
608p	문제-본문	$Lift(X \Rightarrow Y) = \frac{Conf(x \Rightarrow y)}{S(Y)} = \frac{\frac{S(X, Y)}{S(X)}}{S(Y)} = \frac{S(X, Y)}{S(X)\,S(Y)} = \frac{\frac{Count(X, Y)}{N}}{\frac{Count(X)}{N}\,\frac{Count(Y)}{N}}$ $= \frac{P(Y \mid X)}{P(Y)} = \frac{\frac{P(Y \mid X)}{P(Y)}}{P(Y)} = \frac{P(X \cap Y)}{P(X)P(Y)}$	$Lift(X \Rightarrow Y) = \frac{Conf(X \Rightarrow Y)}{S(Y)} = \frac{\frac{S(X, Y)}{S(X)}}{S(Y)} = \frac{S(X, Y)}{S(X)S(Y)} = \frac{\frac{Count(X, Y)}{N}}{\frac{Count(X)}{N}\,\frac{Count(Y)}{N}}$ $= \frac{P(Y \mid X)}{P(Y)} = \frac{\frac{P(X \cap Y)}{P(X)}}{P(Y)} = \frac{P(X \cap Y)}{P(X)P(Y)}$
710p 번호 : 22	문제-문항	① 가장 약한 상관관계는 상관계수가 0이다.	① 상관계수가 0이면 서로 상관관계가 없다.

도서의 오류로 학습에 불편드린 점 진심으로 사과드립니다.
더 나은 도서를 만들기 위해 노력하는 시대교육그룹이 되겠습니다.