[p.33]

예를 들어 두 개의 행렬 $A=\begin{bmatrix}19 & 22\\43 & 50\end{bmatrix}$과 $B=\begin{bmatrix}5 & 6\\7 & 8\end{bmatrix}$이면

$(AB)=\begin{bmatrix}19 & 22\\43 & 50\end{bmatrix}\begin{bmatrix}5 & 6\\7 & 8\end{bmatrix}=\begin{bmatrix}95 & 132\\301 & 400\end{bmatrix}$이고, $(AB)^T=\begin{bmatrix}95 & 301\\132 & 400\end{bmatrix}$

이제 $A^T=\begin{bmatrix}19 & 43\\22 & 50\end{bmatrix}$과 $B^T=\begin{bmatrix}5 & 7\\6 & 8\end{bmatrix}$을 곱해보면 다음과 같다.

$$B^TA^T=\begin{bmatrix}5 & 7\\6 & 8\end{bmatrix}\begin{bmatrix}19 & 43\\22 & 50\end{bmatrix}=\begin{bmatrix}95 & 301\\132 & 400\end{bmatrix}$$

따라서 $(AB)^T=B^TA^T$가 성립한다.

〉〉

예를 들어 두 개의 행렬 $A=\begin{bmatrix}19 & 22\\43 & 50\end{bmatrix}$과 $B=\begin{bmatrix}5 & 6\\7 & 8\end{bmatrix}$이면

$(AB)=\begin{bmatrix}19 & 22\\43 & 50\end{bmatrix}\begin{bmatrix}5 & 6\\7 & 8\end{bmatrix}=\begin{bmatrix}249 & 290\\565 & 658\end{bmatrix}$이고, $(AB)^T=\begin{bmatrix}249 & 565\\290 & 568\end{bmatrix}$

이제 $A^T=\begin{bmatrix}19 & 43\\22 & 50\end{bmatrix}$과 $B^T=\begin{bmatrix}5 & 7\\6 & 8\end{bmatrix}$을 곱해보면 다음과 같다.

$$B^TA^T=\begin{bmatrix}5 & 7\\6 & 8\end{bmatrix}\begin{bmatrix}19 & 43\\22 & 50\end{bmatrix}=\begin{bmatrix}249 & 565\\290 & 658\end{bmatrix}$$

따라서 $(AB)^T=B^TA^T$가 성립한다.

A는 n개의 열 벡터인 $c^{(i)} \in \mathbb{R}^{m \times 1}$ $\forall i \in \{1,2,3,\ldots,n\}$으로 구성되며, 다음과 같이 나타낼 수 있다.

$$A = \left[c^{(1)} c^{(2)} c^{(3)} \ldots . c^{(n)} \right]$$

$$b = Ax = \left[c^{(1)} c^{(2)} c^{(3)} \ldots . c^{(n)} \right] \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} = x_1 c^{(1)} + x_2 c^{(2)} + \ldots .. + x_n c^{(n)}$$

위의 행렬과 벡터의 곱은 행렬 A의 열 벡터와 선형 계수^linear coefficient^인 x 벡터의 선형 조합이다. 벡터 x의 구성 요소는 선형 계수다.

곱셈을 통해 형성된 새로운 벡터 b는 A의 열 벡터와 동일한 차원을 가지며 동일한 열 공간에 남는다. 열 벡터를 결합하는 방법에 관계없이 열 벡터가 차지하는 공간을 벗어나지 않는다. 즉 열 벡터 공간과 동일하다.

〉〉

A는 n개의 열^column^을 가진 벡터인 $c^{(i)} \in \mathbb{R}^{m \times 1}$ $\forall i \in \{1,2,3,\ldots,n\}$으로 구성되며, 다음과 같이 나타낼 수 있다.

$$A = \left[c^{(1)} c^{(2)} c^{(3)} \ldots . c^{(n)} \right]$$

$$b = Ax = \left[c^{(1)} c^{(2)} c^{(3)} \ldots . c^{(n)} \right] \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} = x_1 c^{(1)} + x_2 c^{(2)} + \ldots .. + x_n c^{(n)}$$

위의 행렬과 벡터의 곱은 행렬 A의 열 벡터와 선형 계수^linear coefficient^인 x 벡터의 선형 조합이다. 벡터 x의 구성 요소는 선형 계수다.

곱셈을 통해 형성된 새로운 벡터 b는 A와 동일한 차원의 행을 가지며 동일한 개수의 행을 가진다. 행렬 A를 결합하는 방법에 관계없이 행렬 A가 차지하는 공간을 벗어나지 않는다. 즉 A의 행 공간과 동일하다.

[p.35: 2행]

보다시피 A와 $b \in \mathbb{R}^{2\times1}$은 동일한 열 벡터 차원을 갖는다.
–〉
보다시피 A와 $b \in \mathbb{R}^{2\times1}$은 동일한 행 공간을 갖는다.

[p.35: 3행]

벡터가 다른 벡터의 선형 조합으로 표현될 수 있다면 벡터는 다른 벡터에 일차독립이라고 한다.
–〉
벡터가 다른 벡터의 일차 결합으로 표현될 수 있다면 벡터는 다른 벡터에 일차 종속이라고 한다.

[p.35: 5행]

$v_1 = 5v_2 + 7v_3$인 경우에 다른 벡터들의 합으로 표현될 수 있기 때문에 v_1, v_2, v_3 은 일차독립이다.
–〉
$v_1 = 5v_2 + 7v_3$인 경우에 다른 벡터들의 합으로 표현될 수 있기 때문에 v_1, v_2, v_3 은 일차 종속이다.

[p.35: 아래에서 2행]

즉, n개의 벡터를 선형 조합해 n차원 공간에서 모든 가능한 벡터를 생성할 수 있다.
–〉
즉, n개의 벡터를 일차 결합해 n차원 공간에서 모든 가능한 벡터를 생성할 수 있다.

[p.36: 아래에서 2행]

원하는 3차원 벡터를 가져와서 앞서 언급된 세 벡터의 선형 조합으로 표현할 수 있다.
–〉
원하는 3차원 벡터를 가져와서 앞서 언급된 세 벡터의 일차 결합으로 표현할 수 있다.

[p.37: 3행]

벡터 v_1, v_2의 선형 조합인 벡터 v_3을 취했다면, 3차원 공간 전체에 생성할 수 없기 때문에 v_1과 v_2에 생성된 2차원 부분 공간에 국한해야 한다.
–〉
벡터 v_1, v_2의 일차 결합인 벡터 v_3을 취했다면, 3차원 공간 전체에 생성할 수 없기 때문에 v_1과 v_2에 생성된 2차원 부분 공간에 국한해야 한다.

[p.38: 마지막행]

n행 벡터 또는 열 벡터가 일차독립이 아닌 경우,

—〉

n개의 벡터 또는 열 벡터가 일차독립이 아닌 경우,

[p.57: 아래에서 3행]

단변수 함수의 미분값이 0이거나 다변수 함수의 기울기 벡터가 영벡터인 점을 고정점stationary point이라 한다.

—〉

단변수 함수의 미분값이 0이거나 다변수 함수의 기울기 벡터가 영벡터인 점을 정류점stationary point이라 한다.

[p.57: 마지막행]

그림 1-12는 고정점의 종류를 보여준다.

—〉

그림 1-12는 정류점의 종류를 보여준다.

[p.71: 아래에서 2행]

확률 밀도 함수PDF, probability density function는 도메인의 각 값에서 연속 확률 변수의 확률 밀도를 제공한다.

—〉

확률 밀도 함수PDF, probability density function는 정의역의 각 값에서 연속 확률 변수의 확률 밀도를 제공한다.

[p.72: 아래에서 4행]

즉, X는 pmf $P(X=x_i)=p_i$ 를 갖는 이산 확률 변수다.

—〉

즉, X는 확률 질량 함수 $P(X=x_i)=p_i$ 를 갖는 이산 확률 변수다.

[p.73: 5행]

X가 $P(X=x_i)=p_i$ pmf를 가진 n개의 불연속 값을 취하는 이산 확률 변수인 경우, X의 분산은 다음과 같이 나타낼 수 있다.

—〉

X가 $P(X=x_i)=p_i$ 확률 질량 함수를 가진 n개의 불연속 값을 취하는 이산 확률 변수인 경우, X의 분산은 다음과 같이 나타낼 수 있다.

[p.74: 1행]

왜도의 양수 값은 그림 1-20에 설명된 것처럼 데이터의 대량이 왼쪽을 향하는 것을 의미하며, 비대칭의 음의 값은 그림 1-21에 표시된 것처럼 데이터의 대량이 오른쪽을 향한 것을 의미한다.

->

왜도의 양수 값은 그림 1-20에 설명된 것처럼 데이터의 **많은 부분이** 왼쪽을 향하는 것을 의미하며, 비대칭의 음의 값은 그림 1-21에 표시된 것처럼 데이터의 **많은 부분이** 오른쪽을 향한 것을 의미한다.

[p.83: 1행]

또한 다변수 정규분포는 가우스Gaussian 모델의 혼합에서 널리 사용되며,

->

또한 다변수 정규분포는 가우스Gaussian **혼합 모델**에서 널리 사용되며,

[p.83: 3행]

가우스 혼합은 클러스터링, 이상 탐지, 은닉 마르코프 모델 등과 같은 여러 분야에서 사용된다.

->

가우스 **혼합 모델**은 클러스터링, 이상 탐지, 은닉 마르코프 모델 등과 같은 여러 분야에서 사용된다.

[p.87: 2행]

좀 더 명확하게 하기 위해 L이 $p^7(1-p)^3$ 이 될 가능성을 알아보자.

->

좀 더 명확하게 하기 위해 **우도** L이 $p^7(1-p)^3$ 이 될 가능성을 알아보자.

[p.92: 아래에서 7행]

따라서 주택 가격 (y/x')은 입력 벡터 x'에 편향 항 b와 무작위 요소 ϵ의 선형 조합이며 0의 평균과 σ^2의 유한 분산을 갖는 정규분포를 따른다.

->

따라서 주택 가격 (y/x')은 입력 벡터 x'에 편향 항 b와 무작위 요소 ϵ의 **선형결합**이며 0의 평균과 σ^2의 유한 분산을 갖는 정규분포를 따른다.

[p.93: 1행]

m 샘플 $(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), ..., (x^{(m)}, y^{(m)})$이 있다고 가정해보자.

->

m**개의** 샘플 $(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), ..., (x^{(m)}, y^{(m)})$이 있다고 가정해보자.

[p.105~106]

기울기가 음의 방향이 되면, 최소치 쪽으로 움직일 수 있다.

–〉

기울기가 음의 방향이 되면, 최솟값 쪽으로 움직일 수 있다.

[p.107: 수식 1행]

$$C(\theta+\Delta\theta)=C(\theta)+\Delta\theta^T\nabla C(\theta)+\frac{1}{2}\Delta\theta^T H(\theta)\Delta\theta+\textit{higher order terms}$$

–〉

$$C(\theta+\Delta\theta)=C(\theta)+\Delta\theta^T\nabla C(\theta)+\frac{1}{2}\Delta\theta^T H(\theta)\Delta\theta+\text{고차원항}$$

[p.118: 10행]

λ 는 라그랑주 승수법Lagrange multiplier이라 불린다.

–〉

λ 는 라그랑주 승수Lagrange multiplier라 불린다.

[p.130: 1행]

작은 값의 λ 를 취하면 b의 값이 증가하고, u의 노름도 증가하지만 λ 값이 클수록 b가 작아지므로 θ는 더 작아진다.

–〉

작은 값의 λ 를 취하면 b의 값이 증가하고, u의 노름도 증가하지만 λ 값이 클수록 b가 작아지므로 θ의 노름은 더 작아진다.

[p.220: 1행]

N×M차원의 그레이스케일 영상은 공간 좌표의 스칼라 2D 신호로 표현할 수 있다.

–〉

N×M차원의 그레이스케일 이미지는 공간 좌표의 스칼라 2D 신호로 표현할 수 있다.

[p.225: 6행]

좌표 방향곱의 합계 및 임펄스 응답 값을 출력으로 나타낸다.

–〉

좌표 방향끼리의 곱과 임펄스 응답을 합해 출력으로 나타낸다.

[p.233]

예제 3-2에서는 모나리자의 이미지를 읽고 이미지에 가우스 백색 잡음을 도입했다.
–〉
예제 3-2에서는 모나리자의 이미지를 읽고 이미지에 가우스 백색 잡음을 추가했다.

[p.243: 아래에서 7행]

세 번째 계층에서는 이전 계층에서 생성된 기능을 기반으로 훨씬 더 복잡한 기능을 배운다.
–〉
세 번째 계층에서는 이전 계층에서 생성된 기능을 기반으로 훨씬 더 복잡한 특성을 배운다.

[p.246: 4행]

2D 컨볼루션은 공간 차원에서 수행되지만, 이미지 볼륨의 깊이 채널에는 컨볼루션하지 않는다.
–〉
2D 컨볼루션은 공간 차원에서 수행되지만, 이미지 볼륨의 깊이 차원depth dimension에서는 컨볼루션하지 않는다.

[p.286: 아래에서 2행]

ResNet은 깊이라는 개념을 가지고 있으며, 잔여 블록이라는 고유한 아이디어를
–〉
ResNet은 깊이라는 개념을 가지고 있으며, 잔차 블록이라는 고유한 아이디어를

[p.287: 3행]

그러나 ResNet의 잔차 블록 개념을 사용해 입력에서 출력으로의 직접 매핑이 아닌 잔여 매핑을 학습하려고 한다.
–〉
그러나 ResNet의 잔차 블록 개념을 사용해 입력에서 출력으로의 직접 매핑이 아닌 잔차 매핑을 학습하려고 한다.

[p.288: 12행]

초기 계층은 곡선과 가장자리 같은 매우 일반적인 기능을 감지하는 법을 학습한다.
–〉
초기 계층은 곡선과 가장자리 같은 매우 일반적인 특성을 감지하는 법을 학습한다.

[p.303: 마지막행]

구문 분석은 단어가 그룹화돼 문장으로 연결되는 방법을 나타낸다.

—〉

구문 분석은 단어를 그룹화하고 문장으로 연결하는 방법을 나타낸다.

[p.304: 6행]

두 벡터 사이의 코사인을 계산하면, 벡터 구성 간의 유사성을 기반으로 유사성을 측정할 수 있다.

—〉

두 벡터 사이의 코사인을 계산하면, 벡터 간의 유사도를 측정할 수 있다.

[p.307: 아래에서 2행]

정규화 스킴normalizing scheme은 역문서 빈도를 비선형으로 만들기 위해 용어 빈도 n에 적용할 수 있다.

—〉

정규화 스킴normalizing scheme은 역문서 빈도를 비선형으로 만들기 위해 용어 빈도 n을 적용할 수 있다.

[p.308: 1행]

이는 문서 빈도 기여도가 n의 작은 값에 대해 선형이며, n이 증가함에 따라 기여도가 포화된다.

—〉

문서 빈도 기여도는 n이 작은 경우 선형이며, n이 증가함에 따라 기여도가 포화된다.

[p.335: 아래에서 4행]

이는 벡터 간의 차이점을 나타내는 데 가장 좋은 방법인 선형 벡터 공간에서와 같은 방식이다.

—〉

이는 선형 벡터 공간에서도 벡터 간의 차이점을 나타낼 수 있는 가장 좋은 방식이다.

[p.336: 아래에서 4행]

f와 g를 지수 함수exponential function로 선택하면, 확률의 비율이 단어 벡터의 차이를 부호화encode할 수 있고 동시에 동시 발생 확률이 내적에 의존하게 유지할 수 있다.

—〉

f와 g를 지수함수exponential function로 선택하면, 확률의 비율이 단어 벡터의 차이를 인코딩할 수 있고, 그와 함께 동시 발생 확률이 내적에 의존하게 유지할 수 있다.

[p.336: 마지막행]

벡터의 유사성 및 차별의 개념을 유지한다.
->
벡터의 유사성 및 식별력의 개념을 유지한다.

[p.374: 아래에서 6행]

zt에서 업데이트 게이트 출력 단위가 0에 가까워지면
->
zt에서 업데이트 게이트 출력 유닛이 0에 가까워지면

[p.381: 첫행]

여기서 Z는 다음에 의해 주어진 분할 함수partition function다.
->
여기서 Z는 다음에 의해 주어진 분배 함수partition function다.

[p.381: 위에서 5행]

이러한 경우 분할 함수는 다음과 같이 표현될 수 있다.
->
이러한 경우 분배 함수는 다음과 같이 표현될 수 있다.

[p.401: 아래에서 5행]

분할 함수 Z는 계산하기가 어렵고, $P(v, h)$의 계산을 어렵게 만든다. 작은 사건 집합의 경우 분할 함수를 계산할 수 있다.
->
분배 함수 Z는 계산하기가 어렵고, $P(v, h)$의 계산을 어렵게 만든다. 작은 사건 집합의 경우 분배 함수를 계산할 수 있다.

[p.407: 첫행]

분할 함수 Z는 (5.3.20)의 첫 번째 식과 달리
->
분배 함수 Z는 (5.3.20)의 첫 번째 식과 달리