Machine Learning

군집화: 비슷한 데이터들끼리 모음 하지만 분류와 다르게 분류 클래스는 정해지지 않음

밀도 추정: 데이터로부터 변수가 가질 수 있는 모든 값의 확률을 추정

차원 축소: 여러 차원들 중 가장 중요한 몇개의 차원들만 사용

1. 유형

	X(입력값, 특징벡터, input)	Y(출력값, 목표값, output)	
지도학습	○	○	→ 회귀, 분류
비지도학습	○	×	→ 군집화, 밀도 추정, 차원 축소
준지도 학습	○	매우 적음	

강화학습: 에이전트가 한 행동에 대해 환경이 보상과 상태를 반환하고 보상이 증가하는 방향으로 행동 학습

회귀 vs 분류: 회귀는 출력값이 실수, 분류는 클래스

매니폴드 가정: 고차원의 데이터에서 출력값에 영향을 주는 몇몇 중요 특징들이 있다

2. Dataset

훈련집합: 모델 학습을 위한 데이터

검증집합: 학습이 제대로 되었는지 확인하기 위한 데이터

테스트 집합: 실제로 출력값을 예측하기 위한 데이터

통상적으로 데이터 셋의 크기는 특징 공간에 비해 매우 작다. → 데이터를 효과적으로 이용할 필요가 있다.

3. 비용 함수

Ex) MSE(Mean Square Error) / $J(\theta) = \frac{1}{2}\sum_{i=1}^{n}(\underbrace{h_\theta(x^{(i)}) - y^{(i)}}_{\text{오차}})^2$ → 평균 제곱 오차

비용 함수가 linear Model일 경우 해석적인 방법을 이용하면 되나, 고차원일 경우에는 경사 하강법과 같은 반복적인 방법을 통해 최적화를 해야 한다.

Linear Algebra Remind

내적

$$X^T Y = [x_1\ x_2\ \cdots\ x_n]\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i$$

외적

$$X Y^T = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}[y_1\ y_2 \cdots y_n] = \begin{bmatrix} x_1y_1 & x_1y_2 & \\ x_2y_1 & x_2y_2 & \cdots \\ & & x_ny_n \end{bmatrix}$$

1. Tensor(텐서)

3차원 이상의 고차원 수 나열

2. Matrix Multiplication(행렬 곱)

$A \in \mathbb{R}_{m \times n}, B \in \mathbb{R}_{n \times p}$ (곱해지는 행렬의 열과 곱하는 행렬의 행이 같아야 함)

$AB = C \in \mathbb{R}_{m \times p}$, $C_{ij} = \sum_{k=1}^{n} A_{ik} \times B_{kj}$

3. 행렬의 수학적 성질

$A(BC) = (AB)C$

$A(B+C) = AB + AC$ → 결합, 분배 법칙은 성립하지만 교환법칙은 성립하지 않음

$AB \neq BA$

4. 대각합 / 성질

$trA = \sum_{i=1}^{n} A_{ii}$

정사각 행렬 A에 대해 $trA = trA^T$

$tr(A+B) = trA + trB$

$tr(tA) = t\,tr(A)$

정사각행렬 AB에 대해 $tr(AB) = tr(BA)$

정사각행렬 ABC에 대해 $tr(ABC) = tr(BCA) = tr(CAB)$

5. 선형 독립

벡터의 집합 $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\} \subset \mathbb{R}^m$ 에 대해 어떤 벡터가 다른 벡터들의 선형결합으로 표현된다면 선형 종속이라고 하고 그렇지 않다면 선형 독립이라고 한다

↳ $\mathbf{x}_n = \sum_{i=1}^{n-1} a_i \mathbf{x}_i$

6. 공간

벡터의 집합 $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\} \subset \mathbb{R}^m$ 에 대해 벡터들의 선형 결합으로 표현되는 벡터들의 집합을 공간이라고 한다.

7. 역행렬/성질

정사각행렬 A에 대해 $A^{-1}A = AA^{-1} = I$를 만족하는 A^{-1}를 A의 역행렬이라고 하며 이때 A를 invertible 혹은 non-singular라고 한다.

$(A^{-1})^{-1} = A$

$(AB)^{-1} = B^{-1}A^{-1}$

$(A^{-1})^T = (A^T)^{-1} = A^{-T}$

$(A+B)^{-1} \neq A^{-1} + B^{-1}$

선형변환

$\mathbf{v} = A\mathbf{x}$ ($\mathbf{v} \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$, $\mathbf{x} \in \mathbb{R}^n$)

m차원 벡터 v를 n차원 벡터 x로 변환

ex) 회전

$$Rot(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

8. 가우스 소거법

선형 방정식을 첨가행렬로 변환하고 3가지 연산

1. 2개의 행들을 교체한다.
2. 0이 아닌 상수 값을 행에 곱한다.
3. 한 행에 스칼라를 곱해 다른 행에 더한다

을 통해 행 사다리꼴 형태로 만들어서 해를 구하는 방법

ex)
$$\begin{aligned} 2x_1 + x_2 - x_3 &= 8 \\ -3x_1 - x_2 + 2x_3 &= -11 \\ -2x_1 + x_2 + 2x_3 &= -3 \end{aligned} \longrightarrow \underbrace{\begin{bmatrix} 2 & 1 & -1 \\ -3 & -1 & 2 \\ -2 & 1 & 2 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} 8 \\ -11 \\ -3 \end{bmatrix}}_{\mathbf{b}} \longrightarrow \begin{array}{c} \left[\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ -3 & -1 & 2 & -11 \\ -2 & 1 & 2 & -3 \end{array}\right] \end{array} \begin{array}{l} L_1 \\ L_2 \rightarrow 3L_1 + 2L_2 \Rightarrow L_2 \\ L_3 \end{array}$$

[A|b]
첨가행렬

$$\begin{aligned} 2x_1 + x_2 - x_3 = 8 &\rightarrow x_1 = 2 \\ x_2 + x_3 = 2 &\rightarrow x_2 = 3 \\ x_3 &= -1 \end{aligned} \longleftarrow \left[\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & -1 \end{array}\right] \begin{array}{l} \\ L_2' \longrightarrow L_1 + 2L_2' - L_3 \Rightarrow L_3 \\ \end{array}$$

9. Norm: 벡터의 크기

l_1 Norm(맨하튼 놈): $\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$

l_2 Norm(유클리드 놈): $\|\mathbf{x}\|_2 = \sum_{i=1}^{n} \sqrt{x_i^2}$

l_∞ Norm(맥스 놈) $= \max_{i=1,\cdots,n} |x_i|$

l_p Norm(P놈) $= \sum_{i=1}^{n} (|x_i|^p)^{\frac{1}{p}}$

Rank

- Row Rank: 행벡터들의 부분 집합 중 선형 독립인 가장 큰 부분집합의 크기

= Column Rank

= Matrix Rank / if Matrix Rank = 행렬의 크기

↳ Full Rank

10. Frobenius Norm: 행렬의 크기

$$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2}$$

11. 벡터의 각도

$\mathbf{a}, \mathbf{b}$가 이루는 각도 = 코사인 유사도 $= \frac{\mathbf{a}}{\|\mathbf{a}\|} \cdot \frac{\mathbf{b}}{\|\mathbf{b}\|} = \cos\theta$

12. 고유값, 고유벡터

어떤 행렬A 에 대해 $A\mathbf{x} = \lambda\mathbf{x}$ (λ는 스칼라)를 만족하는 벡터 $\mathbf{x}$ 가 존재한다면 λ는 행렬 A의 고유값, $\mathbf{x}$는 λ에 대응하는 고유벡터라고 한다.

13. 고윳값 분해

선형 독립인 행렬 A는 고유벡터를 열벡터로 가지는 행렬 P, 고윳값을 대각성분으로 가지는 대각행렬 Δ에 대해 $A = P\Delta P^{-1}$을 만족한다.

orthogonal vector

$x^T y = 0$, if $\|x\|_2 = \|y\|_2 = 1 \rightarrow$ orthonormal

// Matrix

$U^T U = I = U U^T \Rightarrow U^{-1} = U^T$

직교하면 선형 독립이지만 역은 아님

Projection

$P^2 = P$