Dong-A Univ. (ISPL)

동아대학교
DONG-A UNIVERSITY

Entropy & Decision Tree & KNN Method

컴퓨터공학과

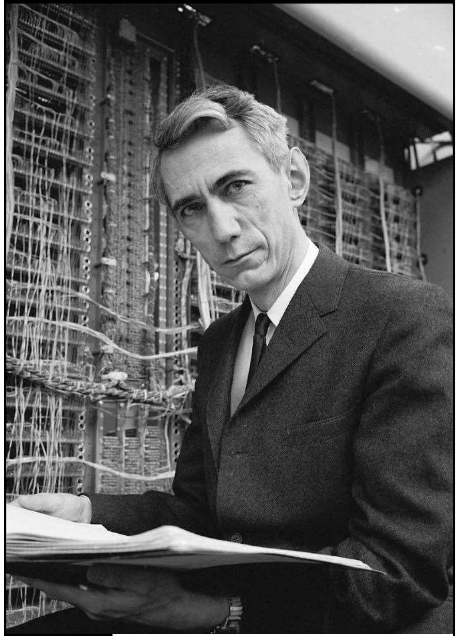
2025년 1학기 머신러닝

Entropy & Decision Tree

- **Entropy: measurement for uncertainty**
  - 불확실성(uncertainty) 정도를 나타내는 수치
  - Shannon entropy: $H(X) = -\sum_{i=1}^{n} p_i \ \log_2 p_i$

Heads of coin	Tails of coin	Calculation	Entropy
50%	50%	-$(0.5 \times \log 0.5 + 0.5 \times \log 0.5)$ = 1	1
100%	0%	-$(1.0 \times \log 1.0 + 0.0 \times \log 0.0)$ = 0	0
90%	10%	-$(0.9 \times \log 0.9 + 0.1 \times \log 0.1)$ = 0	0.47

클로드 섀넌
Claude Shannon



본명	클로드 엘우드 섀넌 Claude Elwood Shannon
출생	1916년 4월 30일 미국 미시간 주 페토스키
사망	2001년 2월 24일 (향년 84세) 미국 매사추세츠 주 메드퍼드
국적	미국
직업	학자, 교수
분야	수학, 컴퓨터과학, 전자공학, 암호학

[펼치기 · 접기]

Entropy & Decision Tree

- **불확실성이 높을수록 Entropy는 큰 값을 가짐**
  - Binary Classification 0 <= Entropy <=1
  - 8-classes Classification 0 <= Entropy <=3
  - 16-classes Classification 0 <= Entropy <=4

Classification Type	Class number (n)	Maximum Entropy for Uniform Distribution H_{max}
Binary	2	$-2 \cdot \frac{1}{2} \log_2 \frac{1}{2} = 1$
8-classes	8	$-8 \cdot \frac{1}{8} \log_2 \frac{1}{8} = 3$
16-classes	16	$-16 \cdot \frac{1}{16} \log_2 \frac{1}{16} = 4$

Entropy & Decision Tree

- **Machine Learning에서 Entropy 활용 예**
  - [1] Deep Learning의 Loss Function
  - [2] Decision Tree
  - [3] Active Learning

- **Machine Learning에서 Entropy 활용 예 (MLP: Loss Function)**
  - [1] Deep Learning의 Loss Function : 학습 모델이 얼마나 잘못 예측하고 있는지는 표현하는 지표
    - 값이 낮을수록 모델이 정확하게 예측했다고 해석할 수 있음
    - Ex. **Cross Entropy Error (CEE)** 계산 방법

$$\mathrm{CEE}(y, y') = -\sum_{i=1}^{N} y_i \times \log(y_i')$$

❖ y: 정답 값

❖ y': 예측 값

$$h(x) = -\sum_{i=1}^{n} (pi \log_2(pi))$$

	0	1	2	3	4	5	6	7	8	9	
	0	0	**1**	0	0	0	0	0	0	0	정답 값 (y, one-hot)
Model A의 예측 결과	0	0	**0.8**	0	0	0	**0.1**	0	**0.1**	0	예측 확률 (y')

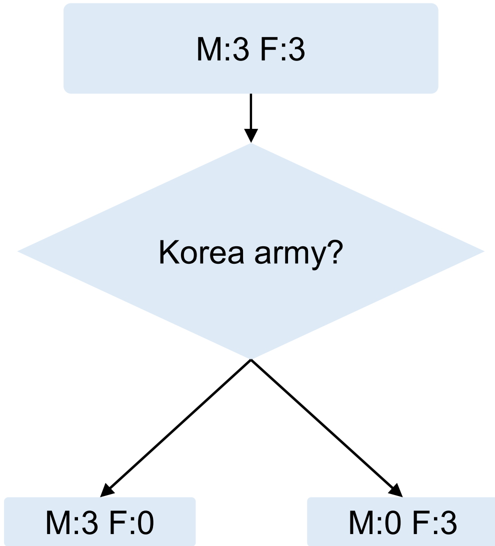
CEE = 0.2231

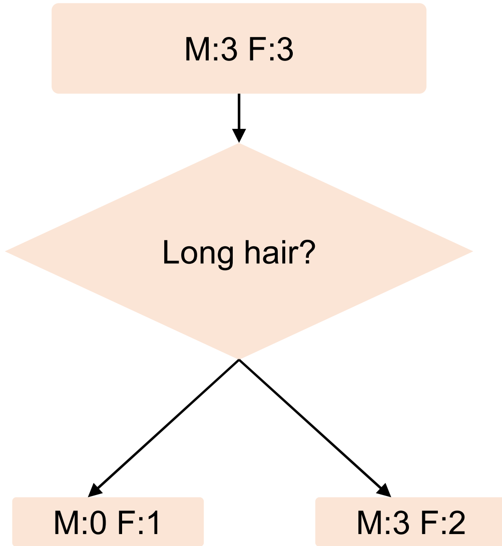
$$\mathrm{CEE}(y, y') = -(1 \times \log(0.8)) = 0.2231$$

Entropy & Decision Tree

- **Machine Learning에서 Entropy 활용 예 (Decision Tree, DT)**
  - [2] Decision Tree
    - DT에서 **확실히 구분이 되는 특징**을 먼저 구분해 주는 것이 중요
    - 확실히 구분이 되는 특징은 **불확실성(엔트로피)가 작다**는 것을 의미

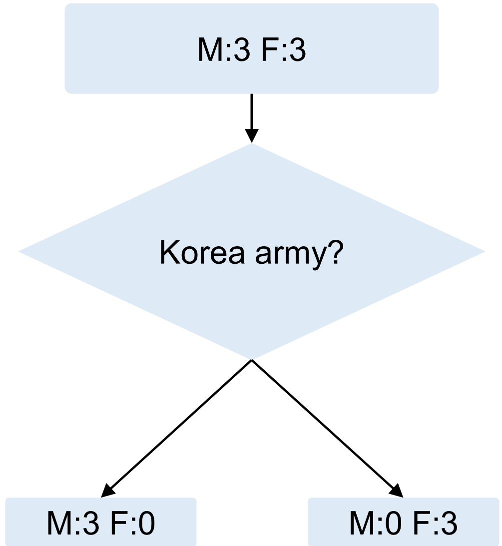
Person	Korea army ?	Long hair?	Gender
1	Yes	No	Male
2	No	No	Female
3	Yes	No	Male
4	No	Yes	Female
5	Yes	No	Male
6	No	No	Female



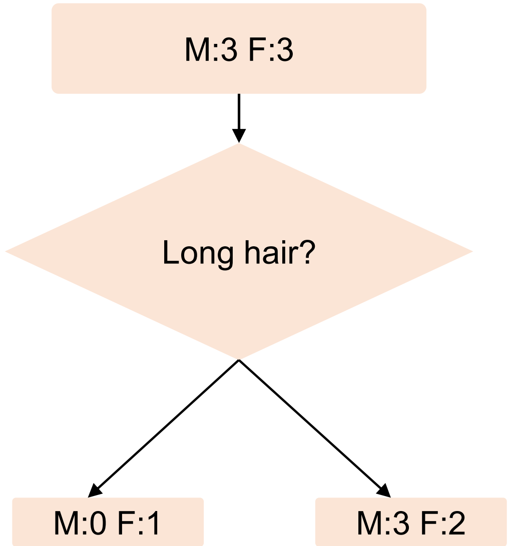


Entropy & Decision Tree

- **Machine Learning에서 Entropy 활용 예 (Decision Tree, DT)**
  - [2] Decision Tree
    - ➢ DT에서 **확실히 구분이 되는 특징**을 먼저 구분해 주는 것이 중요
    - ➢ 확실히 구분이 되는 특징은 **불확실성(엔트로피)가 작다**는 것을 의미



$$\boldsymbol{h(x)} = -\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} (\boldsymbol{pi}\ \mathbf{log_2}(\boldsymbol{pi}))$$



$$\frac{3}{6} * \left(\left(-\frac{3}{3}\right) * \log\left(\frac{3}{3}\right) - \left(-\frac{0}{3}\right) * \log\left(\frac{0}{3}\right)\right) + \frac{3}{6} * \left(\left(-\frac{3}{3}\right) * \log\left(\frac{3}{3}\right) - \left(-\frac{0}{3}\right) * \log\left(\frac{0}{3}\right)\right)$$

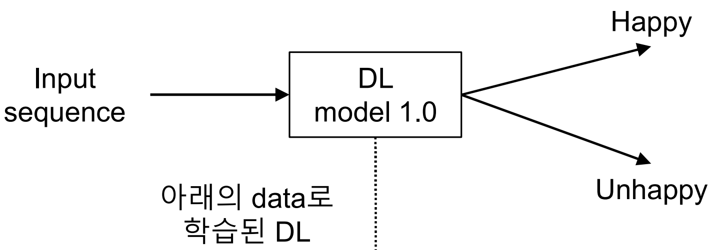
$$totoal\ Entropy : \mathbf{0}$$

$$\frac{1}{6} * \left(\left(-\frac{0}{1}\right) * \log\left(\frac{1}{0}\right) - \left(-\frac{1}{1}\right) * \log\left(\frac{1}{1}\right)\right) + \frac{5}{6} * \left(\left(-\frac{3}{5}\right) * \log\left(\frac{3}{5}\right) - \left(-\frac{2}{5}\right) * \log\left(\frac{2}{5}\right)\right)$$

$$totoal\ Entropy : \mathbf{0.966}$$

Entropy & Decision Tree

- **Machine Learning에서 Entropy 활용 예**
  - [3] Active Learning
    - $h(x) = -\sum_{i=1}^{n}(pi\ \log_2(pi))$

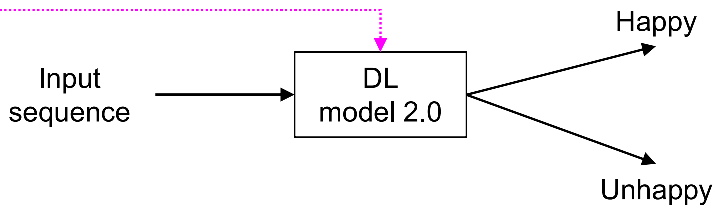


	Happy	Unhappy	Entropy
I love you	0.99	0.01	0.08
I am angry	0.05	0.95	0.29
I am sad	0.3	0.7	0.88
I am feeing blue	0.6	0.4	0.97

uncertain

certain

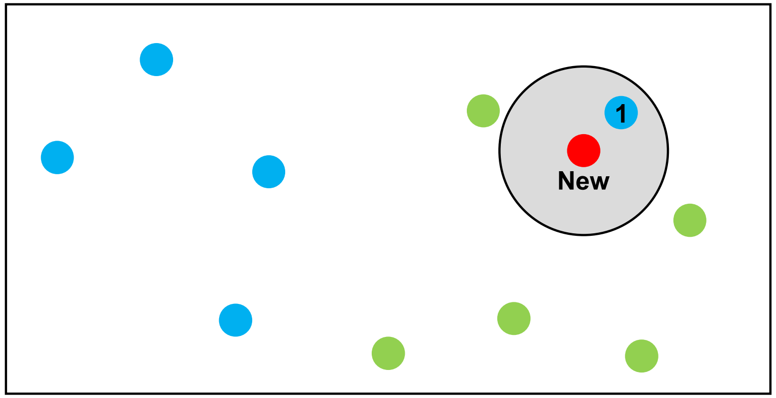
retrain



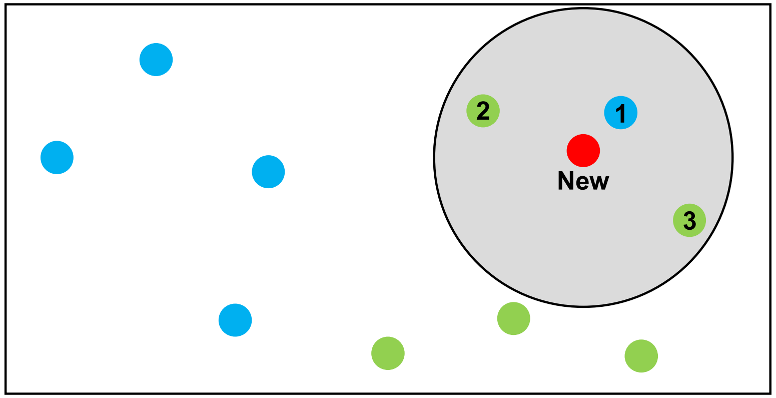
	Happy	Unhappy
I love you	0.99	0.01
I am angry	0.05	0.95
I am sad	0.01	0.99
I am feeing blue	0.04	0.96

K-Nearest Neighbor (KNN)

- **Supervised Learning: <u>Model-based Learning</u>**
  - Linear/Ridge/Lasso/Elastic Regression
  - Deep Learning(MLP & CNN)
  - Support Vector Machine
  - Decision Tree
  - KNN
- **Unsupervised Learning**
  - K-means Algorithm): [Memory-based Learning] or [Lazy Learning]



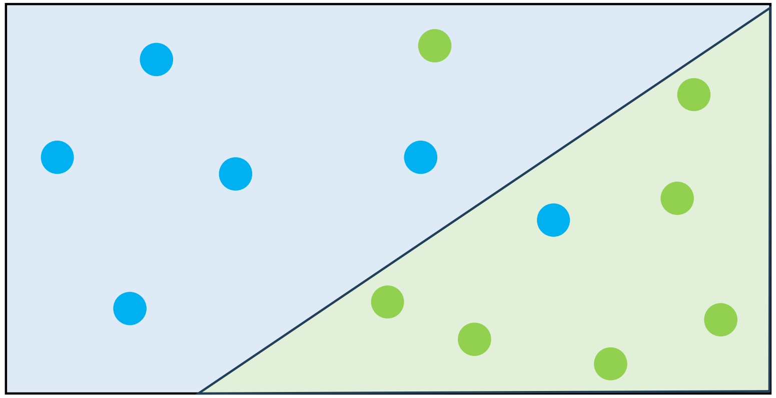
K = 1 예시



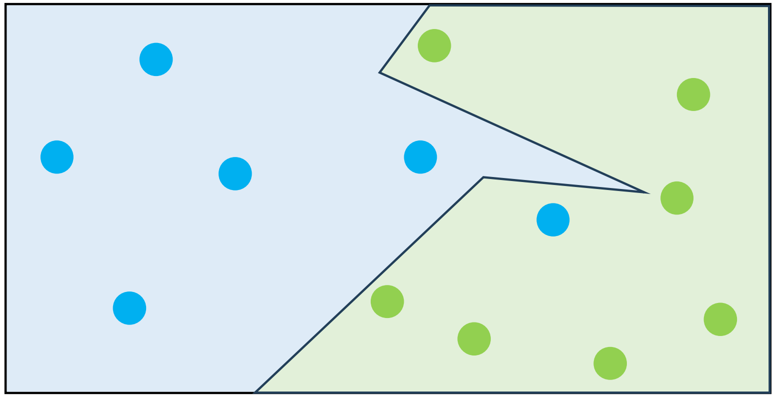
K = 3 예시

K-Nearest Neighbor (KNN)

- **KNN Algorithm**
  - Linear vs non Linear



K = 1 (Linear boundary)

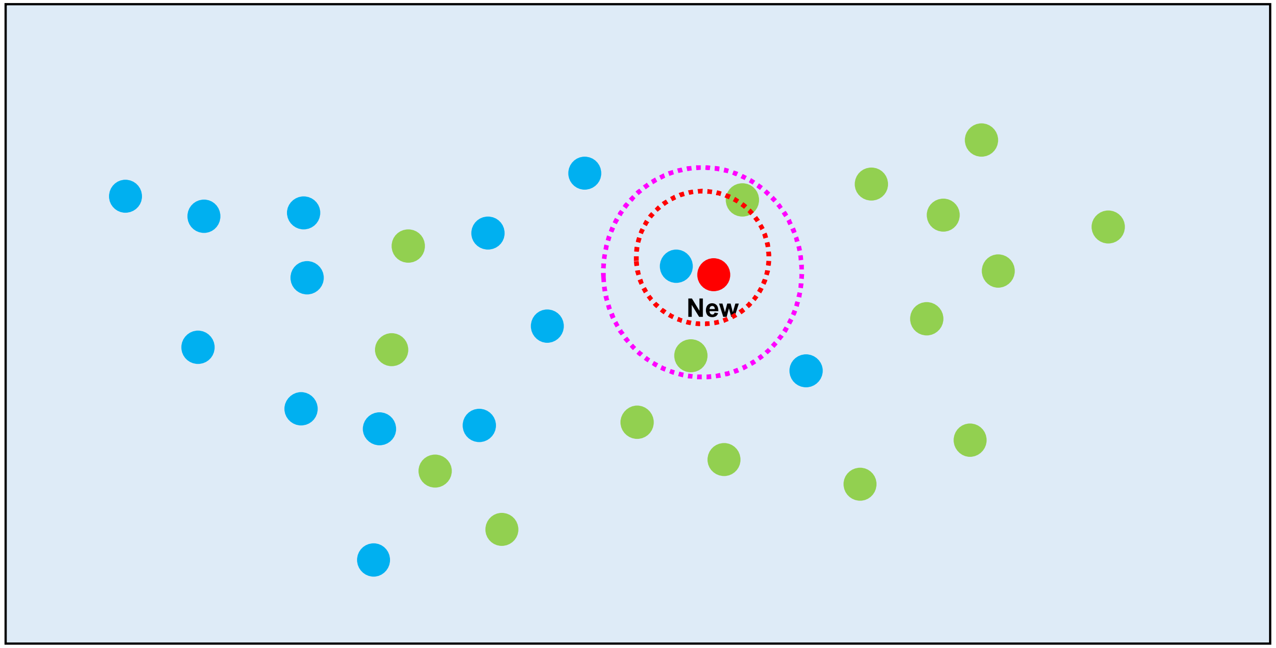


K = 3 (non Linear boundary)

  - **KNN 응용: (1) KNN 분류, (2) KNN 추정**

K-Nearest Neighbor (KNN)

- **KNN 분류**
  - 인접한 K개의 data로부터 Majority voting
    - ➢ K = nearest neighbors
    - ➢ K = 1경우 : 빨간 점선
    - ➢ K = 3경우 : 핑크 점선



K-Nearest Neighbor (KNN)

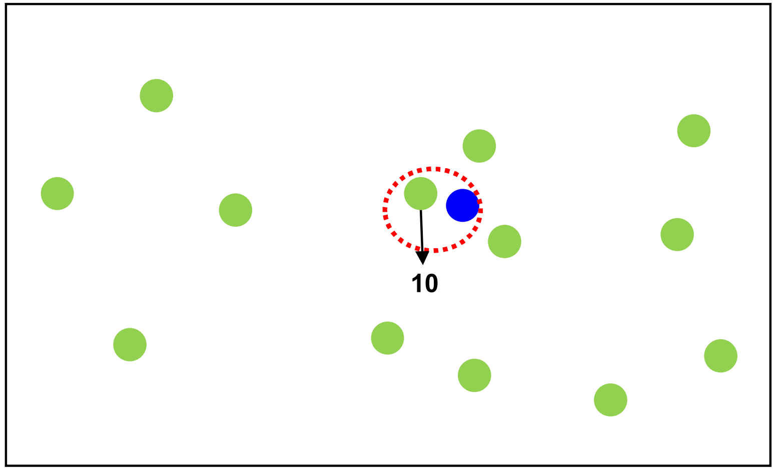
- **KNN 분류**
  - 인접한 K개의 data로부터 Majority voting

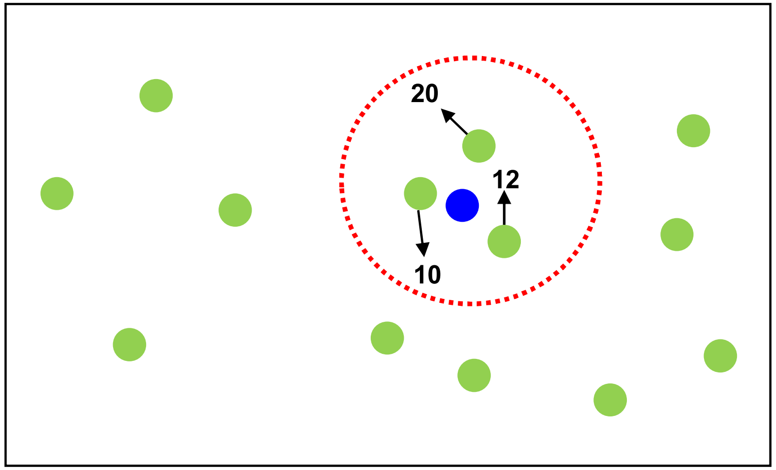
/	감기에 대한 정보				환자 상태	/
사람	**기침**	**콧물**	**가래**	**발열**	**감기 유무**	**새로운 관측치와 거리**
A	1.85	3.4	4.12	2.95	정상	1.54
B	2.9	3.2	3.77	3.1	정상	0.76
C	2.35	2.95	5.25	3.48	정상	2.00
D	3.7	3.8	4.05	3.85	감기	0.78
E	3.45	2.9	2.95	4.1	감기	1.28
F	3.95	2.6	3.4	4.2	감기	1.31
G	**3.05**	**3.1**	**3.95**	**3.7**	**?**	

K = 1 일 때, 정상
K = 3 일 때, 감기

K-Nearest Neighbor (KNN)

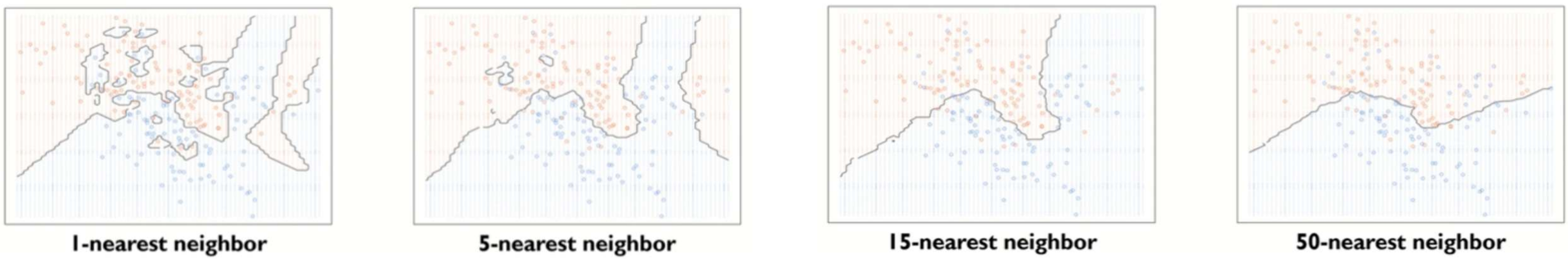
- **KNN 분류**
  - 인접한 K개의 data로부터 평균/중간 값/Min/Max 중 택
    - K = number of nearest neighbors
    - K = 1 : new =15
    - K = 3 : new = (10+12+20)/3 = 14





K-Nearest Neighbor (KNN)

- **KNN Algorithm 이슈**
  - [1] 최적의 K 를 어떻게 결정할 것인가? → 인접한 학습 data를 몇 개까지 탐색할 것인가?
    (1<= K <= 전체 data 개수 → Overfitting vs Underfitting)



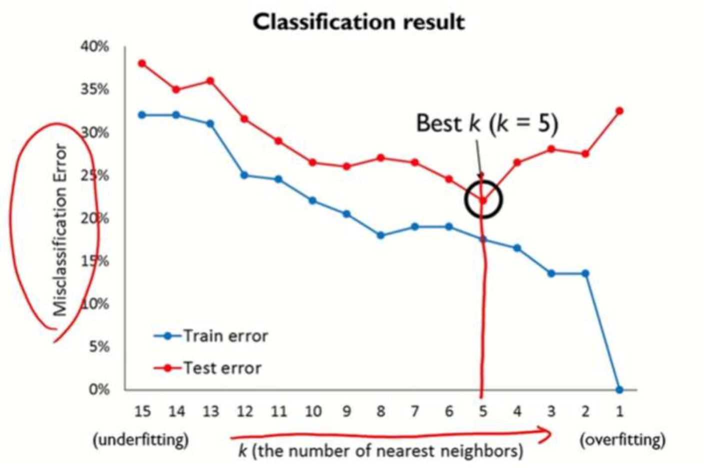
Ref:https://velog.io/@baksugalchae/K-nearest-neighbors-Distance-Measures

K-Nearest Neighbor (KNN)

- **KNN Algorithm 이슈**
  - [1] 최적의 K 를 어떻게 결정할 것인가? → 인접한 학습 data를 몇 개까지 탐색할 것인가?
    - 분류모델: $MisclassError_k = \frac{1}{k}\sum_{i=1}^{k} I(c_i \neq \hat{c_i})\ for\ k = 1,2,\ldots,k^*$

      $I(\cdot): Indicator\ Funcion$

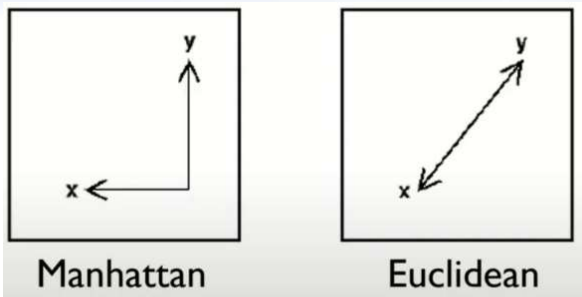
    - $SSE_k = \sum_{i=1}^{k}(y_i - \hat{y_i})^2\ for\ k = 1,2,\ldots,k^*$



Ref:https://velog.io/@baksugalchae/K-nearest-neighbors-Distance-Measures

K-Nearest Neighbor (KNN)

- **KNN Algorithm 이슈**
  - [2] Data간 거리는 어떻게 측정할 것인가? → Distance Measurements

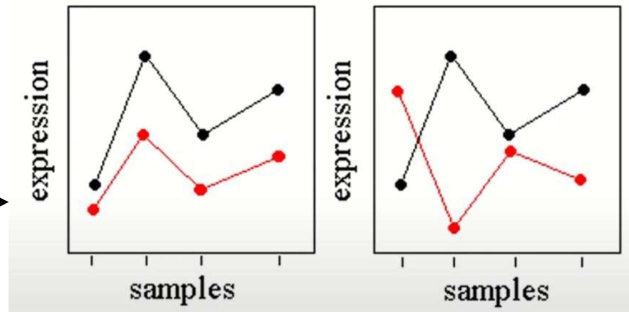


➢ L1 Norm (Manhattan Distance): $D_{Manhattan(X,Y)} = \sum_{i=1}^{n} |x_i - y_i|$

➢ L2 Norm (Euclidean Distance): $d_{(A,B)} = \sqrt{(a_1 - b_1)^2 + \cdots + \left(a_p - b_p\right)^2} = \sqrt{\sum_{i=1}^{p}(a_i - b_i)^2}$

➢ Mahalanobis: $d_{Mahalanobis\,(X,Y)} = \sqrt{(X-Y)^T \Sigma^{-1}(X\, -Y)}$, Σ^{-1} $: inverse\ of\ convariance\ matrix$

➢ Correlation Distance: $d_{corr(X,Y)} = 1 - r\ , where\ =\ \sigma_{XY}$



Ref:https://velog.io/@baksugalchae/K-nearest-neighbors-Distance-Measures

Questions & Answers

Dongsan Jun (dsjun@dau.ac.kr)
Image Signal Processing Laboratory (www.donga-ispl.kr)
Dong-A University, Busan, Rep. of Korea