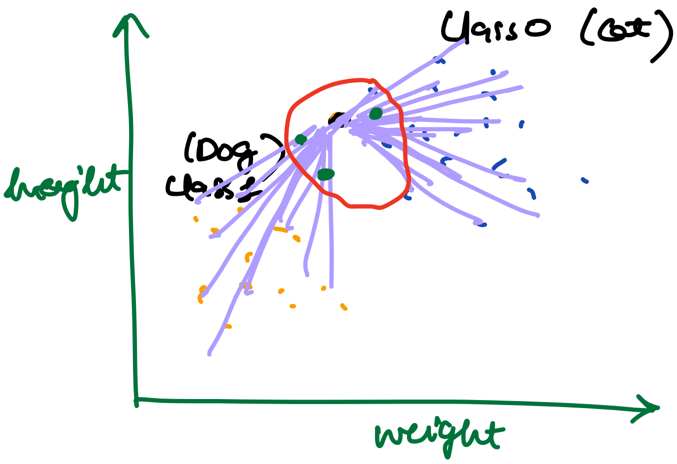
KNN (K Nearest Algorithm)

→ Classification & Regression
   $y \in$ discrete $y \in$ continuous

→ Supervised ML Algo



① Distances b/w test point & all the points

—, —, —, —, —, — , —

② Distance sort inc order

— small
—
—
—
—
— larger

③ $\underline{\underline{k}}$ given

Sorted list → top k values

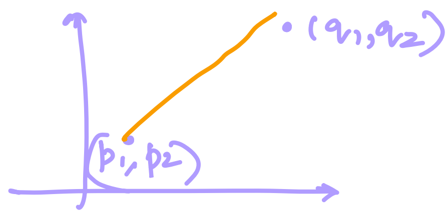
④ P1 → 1
P2 → 1
P3 → 0

majority vote

⑤ test point → class 1.

Distance

① Euclidean Distance

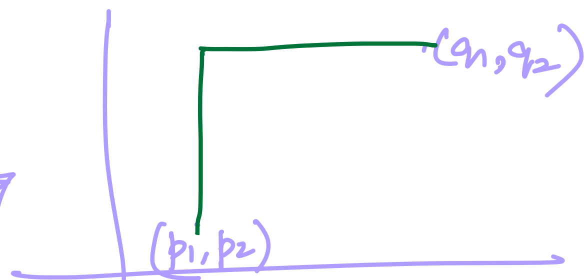


$$\sqrt{(p_1-q_1)^2+(p_2-q_2)^2}$$

$$\left(\sum_{i=1}^{n}(p_i-q_i)^2\right)^{\frac{1}{2}}$$

k=2

② Manhattan Distance



k=1

$$(p_1-q_1)+(p_2-q_2)$$

$$\sum_{i=1}^{n}(p_i-q_i)$$

Minkowski Distance

$$\left(\sum_{i=1}^{n}|p_i-q_i|^k\right)^{\frac{1}{k}}$$

Training time: $O(1)$

Test Time:

1 test point: $N + N\log N + k$

N → Distance, $N\log N$ → Sorting

q test point: $q(N + N\log N + k)$

Q:

Paper → Good / Bad   y

Paper → Durability, Strength

$\sqrt{(7-5)^2+(7-5)^2}$

D	S	labels	Distance	k=1	k=3
7	7	G	$2\sqrt{2}$		
6	4	B	$\sqrt{2}$		✓ B
7	4	G	$\sqrt{5}$		
6	5	G	1	Good ✓	✓ G
3	4	B	$\sqrt{5}$		
1	4	B	$\sqrt{17}$		
4	3	B	$\sqrt{5}$		
3	5	B	2		✓ B

k=1
k=3

(5,5)?

Bad

Naive Bayes Classifier

↳ Bayes Theorem

↳ Conditional Probability

Conditional Probability:

$$P(A|B) = \frac{P(A\cap B)}{P(B)}$$

Eg: family 2 children

already known 1 child is Girl.

Probability that both are girls?

Way 1:

$$S = \{ \cancel{BB}, \underbrace{BG, GB, GG}_{\frac{1}{3}} \}$$

$$P(GG) = 1/3$$

Way 2:

$E \rightarrow$ One is girl $\rightarrow \{GG, BG, GB\}$

$F \rightarrow$ both are girls $\rightarrow \{GG\}$

$$P(F|E) = \frac{P(F\cap E)}{P(E)} = \frac{1/4}{3/4} = 1/3$$

Formulas:

$$P(A|B) = \frac{P(A\cap B)}{P(B)} \Rightarrow P(A\cap B) = P(A|B) \cdot P(B)$$

$$P(B|A) = \frac{P(B\cap A)}{P(A)} \Rightarrow P(B\cap A) = P(B|A) \cdot P(A)$$

$$P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{P(B\cap A)}{P(B)} = \frac{P(B|A) \cdot P(A)}{P(B)}$$

$$\boxed{P(A|B) = \frac{P(B|A)\cdot P(A)}{P(B)}}$$ Bayes Theorem

eg:

A	B	C
2R	1R	3R
3G	4G	2G

Q: Prob. of getting a red ball given that A box is chosen.

$$P(R|A) = 2/5$$

Q: Prob. of getting a red ball?

$$P(R) = P(R\cap A) + P(R\cap B) + P(R\cap C)$$

Q: Prob. that bag A is chosen given that Red ball is drawn?

$$P(A|R) = \frac{P(A\cap R)}{P(R)}$$

$$= \frac{\overset{2/5}{P(R|A)}\ \overset{1/3}{P(A)}}{P(R\cap A) + P(R\cap B) + P(R\cap C)}$$

Naive Bayes Classifier:

$$P(A|B) = \frac{P(B|A)\ P(A)}{P(B)}$$

P(A) → Prior Probability

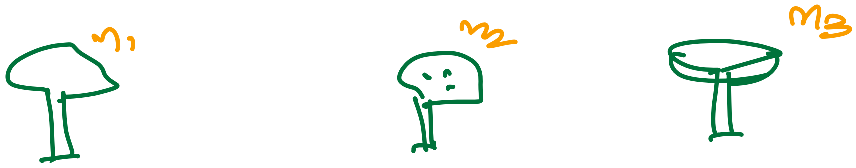
P(B) → prob. of B

P(A|B) → Posterior probability

Conditional prob. of B given A
(Likelihood)

Prior Probability

mushroom dataset:



Class 1 2 3

$$P(y=1) = \frac{m_1}{m_1+m_2+m_3}$$

$$P(y=2) = \frac{m_2}{m_1+m_2+m_3}$$

$$P(y=3) = \frac{m_3}{m_1+m_2+m_3}$$

features: Shape, color, radius, weight . . . .

(a) Test Mushroom —, —, —, —. class $\rightarrow$ 1, 2, 3 ?

Ans?

$P(y=1|x) \rightarrow 0.25$

$P(y=2|x) \rightarrow 0.15$

$P(y=3|x) \rightarrow 0.6 \longrightarrow$ Belongs to Class 3

Sum of all these prob. should be 1

$$P(y=1|x) = \frac{P(x|y=1) * P(y=1)}{P(x)}$$

$$= \frac{P(x|y=1) * P(y=1)}{\underbrace{P(x|y=1).P(y=1)}_{P(x \cap y=1)} + \underbrace{P(x|y=2).P(y=2)}_{P(x \cap y=2)} + \underbrace{P(x|y=3).P(y=3)}_{P(x \cap y=3)}}$$

$$P(y=2|x) = \frac{P(x|y=2) * P(y=2)}{P(x|y=1).P(y=1) + P(x|y=2).P(y=2) + P(x|y=3).P(y=3)}$$

$$P(y=3 \mid x) = \frac{P(x \mid y=3) * P(y=3)}{P(x \mid y=1) \cdot P(y=1) + P(x \mid y=2) \cdot P(y=2) + P(x \mid y=3) \cdot P(y=3)}$$

$$P(y=1 \mid x) = \frac{P(x \mid y=1) \cdot P(y=1)}{\text{Denom}}$$

$$P(y=2 \mid x) = \frac{P(x \mid y=2) \cdot P(y=2)}{\text{Denom}}$$

$$P(y=3 \mid x) = \frac{P(x \mid y=3) \cdot P(y=3)}{\text{Denom}}$$

Ignore Denom

$$P(y=1 \mid x) \propto \boxed{\overset{?}{P(x \mid y=1)}} \cdot P(y=1)$$

$$P(y=2 \mid x) \propto P(x \mid y=2) \cdot P(y=2)$$

$$P(y=3 \mid x) \propto P(x \mid y=3) \cdot P(y=3)$$

$$P(x \mid y=1) = P(x_1 \mid y=1) \cdot P(x_2 \mid y=1) \cdot P(x_3 \mid y=1) \cdot \ldots \cdot P(x_n \mid y=1)$$

x data point

multiple features

independent

$$P(y=1 \mid x) \propto \prod_{i=1}^{n} P(x_i \mid y=1) \cdot P(y=1)$$

$y \in C$

$$P(y=c \mid x) \propto \prod_{i=1}^{n} \underbrace{P(x_i \mid y=c)}_{\text{conditional prob}} \cdot P(y=c)$$

Posterior prob

Likelihood

Prior

SVM

PCA

DT/Forest

Network