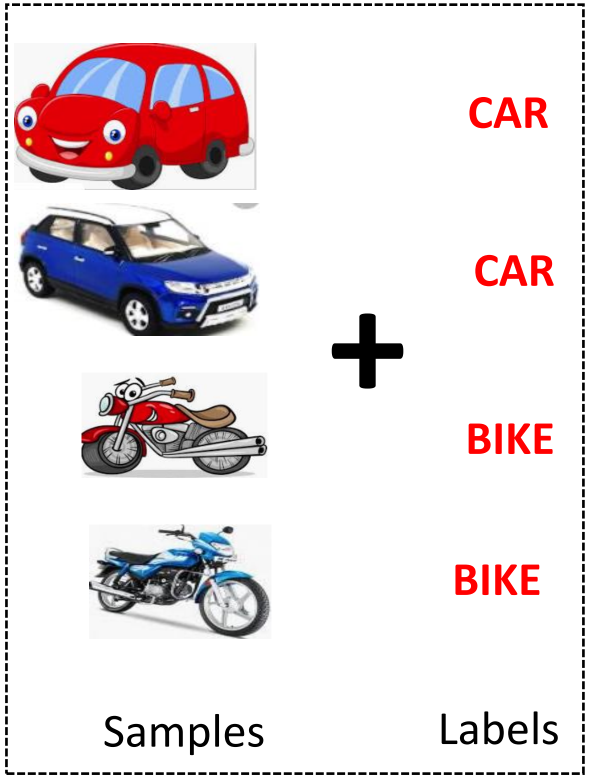
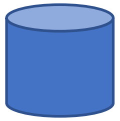
Lesson 3

Few Classification Techniques

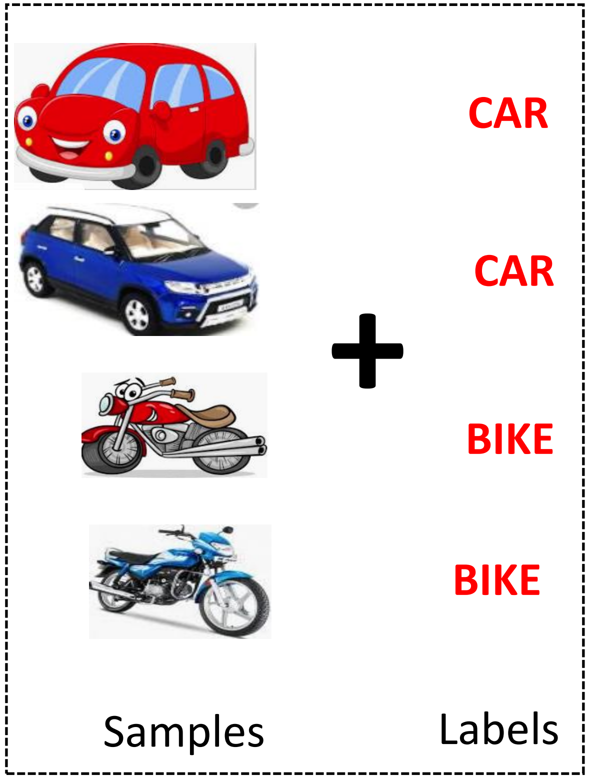
What is Classification?



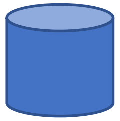
Training Dataset

$f($  ,  $)=$ CAR/BIKE

Classification



Training Dataset

f ( , ) = CAR/BIKE

Given a dataset $\boldsymbol{D = \{ x_1, x_2\, x_3 \dots x_n \}}$ and set of class labels $\boldsymbol{C = \{ c_1\, c_2\, c_3 \dots c_k \}}$, the task of classification to devise a mapping function $\boldsymbol{f : D \rightarrow C}$.

Classification

- Bayesian Classifier
- K-Nearest Neighbours
- Decision Tree
- Support Vector Machine
- Neural Network

Classification

- Bayesian Classifier
- K-Nearest Neighbours
- Decision Tree
- Support Vector Machine
- Neural Network

Bayesian Classifier

#Wheel	Height	Class Label
4	H	CAR
4	H	CAR
4	H	CAR
2	L	BIKE
2	L	BIKE
2	L	BIKE
4	L	BIKE
2	H	CAR

Pr(CAR| 4,H) = 100%

Pr(BIKE| 4,L) = 100%

Pr(CAR| 2,H) = 100%

Pr(BIKE| 2,L) = 100%

Pr(CAR| 4,L) = 0%

Pr(BIKE|4,H) = 0%

Pr(CAR| 2,L) = 0%

Pr(BIKE| 2,H) = 0%

{2 H} ?

$Pr(c_i|x), \forall c_i \epsilon C$

$$\boldsymbol{class} = \arg\max_{c_i} \Pr(c_i|x)$$

Bayesian Classifier

	#Wheel	Height	Class Label
	4	H	CAR
	4	H	CAR
	4	H	CAR
	2	L	BIKE
	2	L	BIKE
	2	L	BIKE
	4	L	BIKE
	2	H	CAR

Pr(CAR| 4,H) = 100%

Pr(BIKE| 4,L) = 100%

Pr(CAR| 2,H) = 100%

Pr(BIKE| 2,L) = 100%

Pr(CAR| 4,L) = 0%

Pr(BIKE|4,H) = 0%

Pr(CAR| 2,L) = 0%

Pr(BIKE| 2,H) = 0%

 {2 H} ?

$$Pr(c_i|x), \forall c_i \epsilon C$$

$$\textbf{class} = \underset{c_i}{\arg\max} \Pr(c_i|x)$$

$Pr(CAR \mid$  $)$

$$Pr(CAR\,|\{2, H\}) = 1$$

$$Pr(BIKE\,|\{2, H\}) = 0$$

class = CAR

Bayes Rule

$\Pr(c_i|x)$

Bayes Rule

$$\Pr(c_i|x) \quad = \frac{\Pr(c_i, x)}{\Pr(x)}$$

Bayes Rule

$$\Pr(c_i|x) \quad = \frac{\Pr(c_i, x)}{\Pr(x)} \quad = \frac{\Pr(x|c_i)\Pr(c_i)}{\Pr(x)}$$

Bayes Rule

$$\Pr(c_i|x) = \frac{\Pr(c_i, x)}{\Pr(x)} = \frac{\Pr(x|c_i)\Pr(c_i)}{\Pr(x)}$$

#Wheel	Height	Class Label
4	H	CAR
4	H	CAR
4	H	CAR
2	L	BIKE
2	L	BIKE
2	L	BIKE
4	L	BIKE
2	H	CAR

Bayes Rule

$$\Pr(c_i|x) = \frac{\Pr(c_i, x)}{\Pr(x)} = \frac{\Pr(x|c_i)\Pr(c_i)}{\Pr(x)}$$

	#Wheel	Height	Class Label
	4	H	CAR
	4	H	CAR
	4	H	CAR
	2	L	BIKE
	2	L	BIKE
	2	L	BIKE
	4	L	BIKE
	2	H	CAR

c_i

Bayes Rule

Prior

$$\Pr(c_i|x) = \frac{\Pr(c_i, x)}{\Pr(x)} = \frac{\Pr(x|c_i)\Pr(c_i)}{\Pr(x)}$$

#Wheel	Height	Class Label
4	H	CAR
4	H	CAR
4	H	CAR
2	L	BIKE
2	L	BIKE
2	L	BIKE
4	L	BIKE
2	H	$\boxed{\text{CAR}}$

c_i

Bayes Rule

$$\Pr(c_i|x) = \frac{\Pr(c_i, x)}{\Pr(x)} = \frac{\Pr(x|c_i)\,\Pr(c_i)}{\Pr(x)}$$

	#Wheel	Height	Class Label
x 	4	H	CAR
	4	H	CAR
	4	H	CAR
	2	L	BIKE
	2	L	BIKE
	2	L	BIKE
	4	L	BIKE
	2	H	CAR

Bayes Rule

Likelihood

$$\Pr(c_i|x) = \frac{\Pr(c_i, x)}{\Pr(x)} = \frac{\Pr(x|c_i)\,\Pr(c_i)}{\Pr(x)}$$

x

#Wheel	Height	Class Label
4	H	CAR
4	H	CAR
4	H	CAR
2	L	BIKE
2	L	BIKE
2	L	BIKE
4	L	BIKE
2	H	CAR

Bayes Rule

$$\Pr(c_i|x) = \frac{\Pr(c_i, x)}{\Pr(x)} = \frac{\Pr(x|c_i)\Pr(c_i)}{\Pr(x)}$$

x

	#Wheel	Height	Class Label
	4	H	CAR
	4	H	CAR
	4	H	CAR
	2	L	BIKE
	2	L	BIKE
	2	L	BIKE
	4	L	BIKE
	2	H	CAR

Bayes Rule

$$\Pr(c_i|x) = \frac{\Pr(c_i, x)}{\Pr(x)} = \frac{\Pr(x|c_i)\Pr(c_i)}{\Pr(x)}$$

Evidence

2

	#Wheel	Height	Class Label
	4	H	CAR
	4	H	CAR
	4	H	CAR
	2	L	BIKE
	2	L	BIKE
	2	L	BIKE
	4	L	BIKE
	2	H	CAR

Bayes Rule

$$\Pr(c_i|x) = \frac{\Pr(c_i, x)}{\Pr(x)} = \frac{\Pr(x|c_i)\Pr(c_i)}{\Pr(x)}$$

$$= \frac{\Pr(x|c_i)\Pr(c_i)}{\underbrace{\Pr(x|c_1)\Pr(c_1) + \Pr(x|c_2)\Pr(c_2) + \ \ldots + \Pr(x|c_k)\Pr(c_k)}_{\text{Marginalization}}}$$

2

#Wheel	Height	Class Label
4	H	CAR
4	H	CAR
4	H	CAR
2	L	BIKE
2	L	BIKE
2	L	BIKE
4	L	BIKE
2	H	CAR

Bayes Rule

$$\Pr(c_i|x) = \frac{\Pr(c_i, x)}{\Pr(x)} = \frac{\Pr(x|c_i)\Pr(c_i)}{\Pr(x)}$$

Posterior

CAR

3

	#Wheel	Height	Class Label
	4	H	CAR
	4	H	CAR
	4	H	CAR
	2	L	BIKE
	2	L	BIKE
	2	L	BIKE
	4	L	BIKE
	2	H	CAR

Bayes Rule

$$\Pr(c_i|x) = \frac{\Pr(c_i, x)}{\Pr(x)} = \frac{\Pr(x|c_i)\Pr(c_i)}{\Pr(x)}$$

Posterior

In Bayesian classification, we estimate Posterior of a class given a sample.

→ CAR

2

	#Wheel	Height	Class Label
	4	H	CAR
	4	H	CAR
	4	H	CAR
	2	L	BIKE
	2	L	BIKE
	2	L	BIKE
	4	L	BIKE
	2	H	CAR

Bayesian Classifier

$$\Pr(c_i|x) \quad = \Pr(c_i \mid \{w_1, w_2 w_3 \dots w_k\}) \; = \frac{\Pr(\{w_1, w_2 w_3 \dots w_k\}|c_i)\Pr(c_i)}{\Pr(\{w_1, w_2 w_3 \dots w_k\})}$$

$\Pr(CAR|$  $)$

$$= \Pr(CAR \mid \{4, H\}) = \frac{\Pr(\{4,H\}|CAR)\Pr(CAR)}{\Pr(\{4, H\})}$$

$$= \frac{0.75 \times 0.5}{0.375} = 1$$

#Wheel	Height	Class Label
4	H	CAR
4	H	CAR
4	H	CAR
2	L	BIKE
2	L	BIKE
2	L	BIKE
4	L	BIKE
2	H	CAR

Bayesian Classifier

$$\Pr(c_i|x) \quad = \Pr(c_i \mid \{w_1, w_2 w_3 \ldots w_k\}) \; = \frac{\Pr(\{w_1, w_2 w_3 \ldots w_k\}|c_i)\Pr(c_i)}{\Pr(\{w_1, w_2 w_3 \ldots w_k\})}$$

$\Pr(CAR|$  $) \quad = \Pr(CAR \mid \{4, H\}) \; = \dfrac{\Pr(\{4, H\}|CAR)\Pr(CAR)}{\Pr(\{4, H\})}$

$$= \frac{0.75 \times 0.5}{0.375} = 1$$

$\Pr(BIKE|$  $) = \Pr(BIKE \mid \{4, H\}) = \dfrac{\Pr(\{4, H\}|BIKE)\Pr(BIKE)}{\Pr(\{4, H\})}$

$$= \frac{0 \times 0.5}{0.375} = 0$$

#Wheel	Height	Class Label
4	H	CAR
4	H	CAR
4	H	CAR
2	L	BIKE
2	L	BIKE
2	L	BIKE
4	L	BIKE
2	H	CAR

Bayesian Classifier

$$\Pr(c_i|x) \;=\Pr(c_i\,|\,\{w_1, w_2 w_3 \ldots w_k\}) \;= \frac{\Pr(\{w_1, w_2 w_3 \ldots w_k\}|c_i)\,\Pr(c_i)}{\Pr(\{w_1, w_2 w_3 \ldots w_k\})}$$

#Wheel	Height	Class Label
4	H	CAR
4	H	CAR
4	H	CAR
2	L	BIKE
2	L	BIKE
2	L	BIKE
4	L	BIKE
2	H	CAR

$\Pr(CAR|$  $)$ $= \Pr(CAR\,|\,\{4,H\}) = \dfrac{\Pr(\{4,H\}|CAR)\,\Pr(CAR)}{\Pr(\{4,H\})}$

$$= \frac{0.75 \times 0.5}{0.375} = 1$$

$\Pr(BIKE|$  $)$ $= \Pr(BIKE\,|\,\{4,H\}) = \dfrac{\Pr(\{4,H\}|BIKE)\,\Pr(BIKE)}{\Pr(\{4,H\})}$

$$= \frac{0 \times 0.5}{0.375} = 0$$

$\boxed{\text{CAR}}$

Bayesian Classifier

$$\Pr(c_i|x) \quad = \Pr(c_i \mid \{w_1, w_2 w_3 \dots w_k\}) \;= \frac{\Pr(\{w_1, w_2 w_3 \dots w_k\}|c_i) \Pr(c_i)}{\Pr(\{w_1, w_2 w_3 \dots w_k\})}$$

$\Pr(CAR|$  $)$

$= \Pr(CAR \mid \{4, H\})$

$$= \frac{\Pr(\{4, H\}|CAR) \Pr(CAR)}{\Pr(\{4, H\})}$$

$\Pr(BIKE|$  $)$

$= \Pr(BIKE \mid \{4, H\})$

$$= \frac{\Pr(\{4, H\}|BIKE) \Pr(BIKE)}{\Pr(\{4, H\})}$$

#Wheel	Height	Class Label
4	H	CAR
4	H	CAR
4	H	CAR
2	L	BIKE
2	L	BIKE
2	L	BIKE
4	L	BIKE
2	H	CAR

Bayesian Classifier

$$\Pr(c_i|x) \quad = \Pr(c_i \mid \{w_1, w_2 w_3 \dots w_k\}) \; = \frac{\Pr(\{w_1, w_2 w_3 \dots w_k\}|c_i) \Pr(c_i)}{\Pr(\{w_1, w_2 w_3 \dots w_k\})}$$

$\Pr(CAR|$  $)$

$= \Pr(CAR \mid \{4, H\})$

$$= \frac{\Pr(\{4, H\}|CAR) \Pr(CAR)}{\Pr(\{4, H\})}$$

$\Pr(BIKE|$  $)$

$= \Pr(BIKE \mid \{4, H\})$

$$= \frac{\Pr(\{4, H\}|BIKE) \Pr(BIKE)}{\Pr(\{4, H\})}$$

Same

#Wheel	Height	Class Label
4	H	CAR
4	H	CAR
4	H	CAR
2	L	BIKE
2	L	BIKE
2	L	BIKE
4	L	BIKE
2	H	CAR

Bayesian Classifier

$$\Pr(c_i|x) \quad = \Pr(c_i \mid \{w_1, w_2 w_3 \ldots w_k\}) \;= \frac{\Pr(\{w_1, w_2 w_3 \ldots w_k\}|c_i)\,\Pr(c_i)}{\Pr(\{w_1, w_2 w_3 \ldots w_k\})}$$

$\Pr(CAR|$  $)$

$= \Pr(CAR \mid \{4, H\})$

$\sim \Pr(\{4, H\}|CAR)\,\Pr(CAR)$

$\Pr(BIKE|$  $)$

$= \Pr(BIKE \mid \{4, H\})$

$\sim \Pr(\{4, H\}|BIKE)\,\Pr(BIKE)$

Relation still maintains

#Wheel	Height	Class Label
4	H	CAR
4	H	CAR
4	H	CAR
2	L	BIKE
2	L	BIKE
2	L	BIKE
4	L	BIKE
2	H	CAR

Bayesian Classifier

$$\Pr(c_i|x) \quad = \Pr(c_i \mid \{w_1, w_2 w_3 \dots w_k\}) \ = \frac{\Pr(\{w_1, w_2 w_3 \dots w_k\}|c_i)\Pr(c_i)}{\Pr(\{w_1, w_2 w_3 \dots w_k\})}$$

$\Pr(CAR|$  $)$

$= \Pr(CAR \mid \{4, H\})$

$\sim \Pr(\{4, H\}|CAR)\Pr(CAR)$

$\Pr(BIKE|$  $)$

$= \Pr(BIKE \mid \{4, H\})$

$\sim \Pr(\{4, H\}|BIKE)\Pr(BIKE)$

Relation still maintains

#Wheel	Height	Class Label
4	H	CAR
4	H	CAR
4	H	CAR
2	L	BIKE
2	L	BIKE
2	L	BIKE
4	L	BIKE
2	H	CAR

Bayesian Classifier

$$\Pr(c_i|x) \;\; = \Pr(c_i \mid \{w_1, w_2, w_3 \ldots w_k\}) = \frac{\Pr(\{w_1, w_2, w_3 \ldots w_k\}|c_i)\,\Pr(c_i)}{\Pr(\{w_1, w_2, w_3 \ldots w_k\})}$$

If ***k*** (the number of classes) is **small,**

#Wheel	Height	Class Label
4	H	CAR
4	H	CAR
4	H	CAR
2	L	BIKE
2	L	BIKE
2	L	BIKE
4	L	BIKE
2	H	CAR

Bayesian Classifier

$$\Pr(c_i|x) \quad = \Pr(c_i \mid \{w_1, w_2, w_3 \ldots w_k\}) = \frac{\Pr(\{w_1, w_2, w_3 \ldots w_k\}|c_i)\Pr(c_i)}{\Pr(\{w_1, w_2, w_3 \ldots w_k\})}$$

If ***k*** (the number of classes) is **small,**

estimating **likelihood** $\Pr(\{w_1, w_2, w_3 \ldots w_k\}|c_i)$ is **feasible**.

#Wheel	Height	Class Label
4	H	CAR
4	H	CAR
4	H	CAR
2	L	BIKE
2	L	BIKE
2	L	BIKE
4	L	BIKE
2	H	CAR

Bayesian Classifier

$$\Pr(c_i|x) \;= \Pr(c_i \mid \{w_1, w_2, w_3 \ldots w_k\}) = \frac{\Pr\left(\{w_1, w_2, w_3 \ldots w_k\}\middle|c_i\right)\Pr(c_i)}{\Pr(\{w_1, w_2, w_3 \ldots w_k\})}$$

However, if ***k*** (the number of classes) is **very large,**

estimating **likelihood** $\Pr\left(\{w_1, w_2, w_3 \ldots w_k\}\middle|c_i\right)$ is **a very expensive task** over **a large dataset.**

Bayesian Classifier

$$\Pr(c_i|x) \quad = \Pr(c_i \mid \{w_1, w_2, w_3 \dots w_k\}) = \frac{\Pr\left(\{w_1, w_{2,} w_3 \dots w_k\}\middle|c_i\right)\Pr(c_i)}{\Pr(\{w_1, w_2, w_3 \dots w_k\})}$$

However, if ***k*** (the number of classes) is **very large,**

estimating **likelihood** $\Pr\left(\{w_1, w_{2,} w_3 \dots w_k\}\middle|c_i\right)$ is **a very expensive task** over **a large dataset.**

$$\Pr\left(\{w_1, w_{2,} w_3 \dots w_k\}\middle|c_i\right) \quad = \Pr(w_1|w_2, w_3 , \dots, w_3, c_i). Pr(w_2|w_3, w_4 , \dots, w_3, c_i) \dots . Pr(w_k|c_i)$$

Bayesian Classifier

$$\Pr(c_i|x) \quad = \Pr(c_i \mid \{w_1, w_2, w_3 \dots w_k\}) = \frac{\Pr(\{w_1, w_2, w_3 \dots w_k\}|c_i)\Pr(c_i)}{\Pr(\{w_1, w_2, w_3 \dots w_k\})}$$

However, if ***k*** (the number of classes) is **very large,**

estimating **likelihood** $\Pr(\{w_1, w_2, w_3 \dots w_k\}|c_i)$ is **a very expensive task** over **a large dataset.**

$$\Pr(\{w_1, w_2, w_3 \dots w_k\}|c_i) \quad = \Pr(w_1|w_2, w_3, \dots, w_3, c_i).Pr(w_2|w_3, w_4, \dots, w_3, c_i) \dots .Pr(w_k|c_i)$$

Naïve Bayes Classifier

$$\Pr(c_i|x) \;\; = \Pr(c_i \mid \{w_1, w_2, w_3 \ldots w_k\}) = \frac{\Pr\left(\{w_1, w_2, w_3 \ldots w_k\} \middle| c_i\right) \Pr(c_i)}{\Pr(\{w_1, w_2, w_3 \ldots w_k\})}$$

To simplify the estimation, we make an **assumption**

- The features are **conditionally independent**.

Naïve Bayes Classifier

$$\Pr(c_i|x) \quad = \Pr(c_i \mid \{w_1, w_2, w_3 \dots w_k\}) = \frac{\Pr\left(\{w_1, w_2, w_3 \dots w_k\}\middle|c_i\right)\Pr(c_i)}{\Pr(\{w_1, w_2, w_3 \dots w_k\})}$$

To simplify the estimation, we make an **assumption**

- The features are **conditionally independent**.

$$\Pr\left(\{w_1, w_2, w_3 \dots w_k\}\middle|c_i\right) \quad = \Pr(w_1|w_2, w_3, \dots, w_3, c_i).\,Pr(w_2|w_3, w_4, \dots, w_3, c_i) \dots .\,Pr(w_k|c_i)$$

Naïve Bayes Classifier

$$\Pr(c_i|x) \;\; = \Pr(c_i \mid \{w_1, w_2, w_3 \ldots w_k\}) = \frac{\Pr\left(\{w_1, w_2, w_3 \ldots w_k\} \middle| c_i\right) \Pr(c_i)}{\Pr(\{w_1, w_2, w_3 \ldots w_k\})}$$

To simplify the estimation, we make an **assumption**

- The features are **conditionally independent**.

$= Pr(w_1|c_i)$ $= Pr(w_2|c_i)$

$$\Pr\left(\{w_1, w_2, w_3 \ldots w_k\} \middle| c_i\right) \;\; = \Pr(w_1|w_2, w_3, \ldots, w_3, c_i)\, Pr(w_2|w_3, w_4, \ldots, w_3, c_i) \ldots Pr(w_k|c_i)$$

Naïve Bayes Classifier

$$\Pr(c_i|x) \quad = \Pr(c_i \mid \{w_1, w_2, w_3 \dots w_k\}) = \frac{\Pr(\{w_1, w_2, w_3 \dots w_k\}|c_i) \Pr(c_i)}{\Pr(\{w_1, w_2, w_3 \dots w_k\})}$$

To simplify the estimation, we make an **assumption**

- The features are **conditionally independent**.

Bayesian: $\Pr(\{w_1, w_2, w_3 \dots w_k\}|c_i) \quad = \Pr(w_1|w_2, w_3, \dots, w_3, c_i). Pr(w_2|w_3, w_4, \dots, w_3, c_i) \dots. Pr(w_k|c_i)$

Naïve Bayes: $\Pr(\{w_1, w_2, w_3 \dots w_k\}|c_i) \quad \sim \Pr(w_1|c_i). Pr(w_2|c_i) \dots. Pr(w_k|c_i) = \prod_{j=1}^{k} \Pr(w_j|c_i)$

Naïve Bayes Classifier

$$\Pr(c_i|x) \quad = \Pr(c_i \mid \{w_1, w_2, w_3 \ldots w_k\}) = \frac{\Pr\left(\{w_1, w_2, w_3 \ldots w_k\} \middle| c_i\right) \Pr(c_i)}{\Pr(\{w_1, w_2, w_3 \ldots w_k\})}$$

$$\sim \prod_{j=1}^{k} \Pr(w_j|c_i) \Pr(c_i)$$

Naïve Bayes Classifier

$$\Pr(c_i|x) \;= \Pr(c_i\,|\,\{w_1, w_2, w_3 \ldots w_k\}) = \frac{\Pr(\{w_1, w_2, w_3 \ldots w_k\}|c_i)\Pr(c_i)}{\Pr(\{w_1, w_2, w_3 \ldots w_k\})}$$

#Wheel	Height	Class Label
4	H	CAR
4	H	CAR
4	H	CAR
2	L	BIKE
2	L	BIKE
2	L	BIKE
4	L	BIKE
2	H	CAR

$$\sim \prod_{j=1}^{k} \Pr(w_j|c_i)\Pr(c_i)$$

$$\Pr(CAR|\,\{4, H\}) = \Pr(4|CAR) \times \Pr(H|CAR) \times \Pr(CAR)$$

$$= 0.75 \times 1 \times 0.5 = 0.375$$

$$\Pr(BIKE|\,\{4, H\}) = \Pr(4|BIKE) \times \Pr(H|BIKE) \times \Pr(BIKE)$$

$$= 0.25 \times 0 \times 0.5 = 0$$

Naïve Bayes Classifier

$$\Pr(c_i|x) = \Pr(c_i \mid \{w_1, w_2, w_3 \dots w_k\}) = \frac{\Pr(\{w_1, w_2, w_3 \dots w_k\}|c_i)\Pr(c_i)}{\Pr(\{w_1, w_2, w_3 \dots w_k\})}$$

#Wheel	Height	Class Label
4	H	CAR
4	H	CAR
4	H	CAR
2	L	BIKE
2	L	BIKE
2	L	BIKE
4	L	BIKE
2	H	CAR

$$\sim \prod_{j=1}^{k} \Pr(w_j|c_i)\Pr(c_i)$$

$$\Pr(CAR|\{4, H\}) = \Pr(4|CAR) \times \Pr(H|CAR) \times \Pr(CAR)$$

$$= 0.75 \times 1 \times 0.5 = 0.375$$

$$\Pr(BIKE|\{4, H\}) = \Pr(4|BIKE) \times \Pr(H|BIKE) \times \Pr(BIKE)$$

$$= 0.25 \times 0 \times 0.5 = 0$$

$\boxed{\text{CAR}}$

What is one of the estimate in the likelihood is zero?

$$\Pr\left(\{w_1, w_2, w_3 \dots w_k\} \middle| c_i\right) \quad \sim \Pr(w_1|c_i).\,Pr(w_2|c_i) \dots .\,Pr(w_k|c_i) = \prod_{j=1}^{k} \Pr(w_j|c_i)$$

$$\Pr(CAR|\,\{4, \boldsymbol{M}\}) = \Pr(4|CAR) \times \Pr(M|CAR) \times \Pr(CAR)$$

$$= 0.75 \times 0 \times 0.5 = 0$$

What is one of the estimate in the likelihood is zero?

$$\Pr\left(\{w_1, w_2, w_3 \ldots w_k\} | c_i\right) \quad \sim \Pr(w_1|c_i).Pr(w_2|c_i)\ \ldots.Pr(w_k|c_i) = \prod_{j=1}^{k} \Pr(w_j|c_i)$$

$$\Pr(CAR|\,\{4, \boldsymbol{M}\}) = \Pr(4|CAR) \times \Pr(M|CAR) \times \Pr(CAR)$$

$$= 0.75 \times 0 \times 0.5 = 0$$

Smoothing

$$Pr(M|CAR) = \frac{0 + Small}{|CAR| + Large}$$

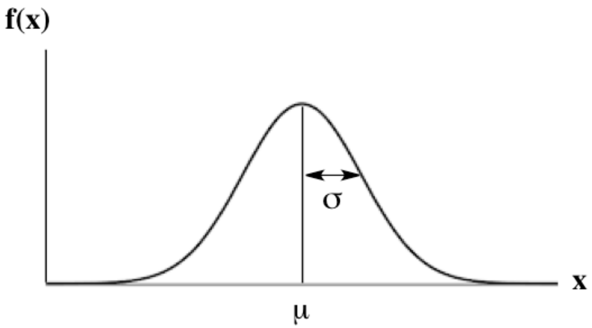
In some of the machine learning tools, you may find

- Naïve Bayes with Gaussian

- Naïve Bayes with Multinomial

In some of the machine learning tools, you may find

- Naïve Bayes with Gaussian



$$f(x) = \frac{1}{\boxed{\sigma}\sqrt{2\pi}} e^{-(x-\boxed{\mu})^2/2\sigma^2}$$

$$e = 2.71828$$

In some of the machine learning tools, you may find

- Naïve Bayes with Multinomial

$$f(x_1,\ldots,x_k;n,p_1,\ldots,p_k) = \Pr(X_1 = x_1 \text{ and } \ldots \text{ and } X_k = x_k)$$

$$= \begin{cases} \dfrac{n!}{x_1!\cdots x_k!}p_1^{x_1}\cdots p_k^{x_k}, & \text{when } \sum_{i=1}^k x_i = n \\ 0 & \text{otherwise,} \end{cases}$$