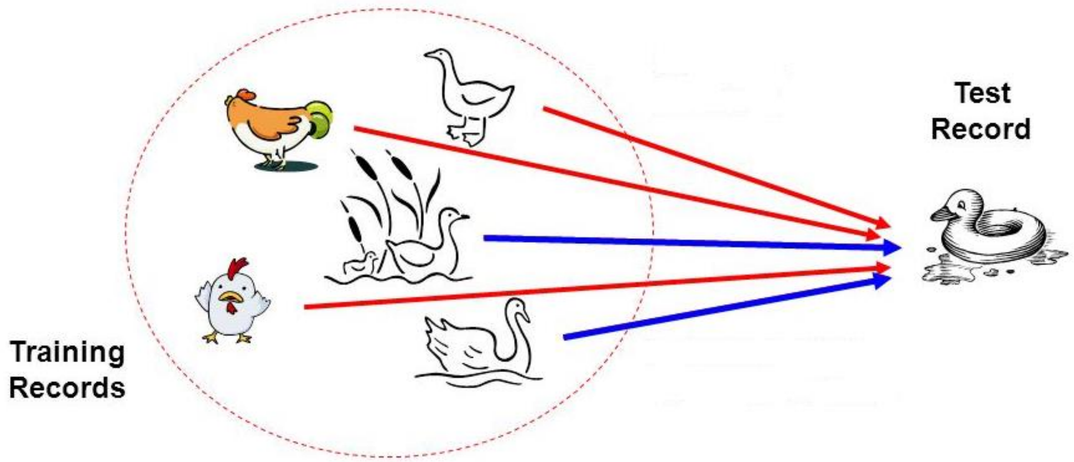
Naïve Bayes Classifier

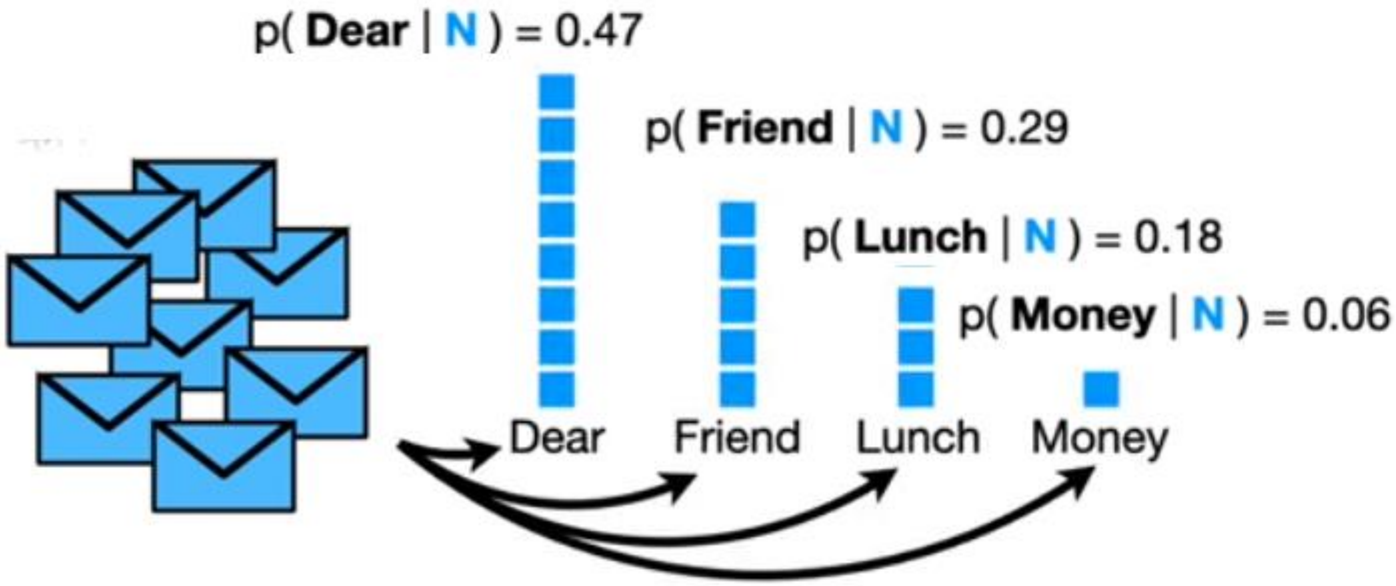
- Given a dataset X = $\{x_1, x_2, \ldots\ldots, x_m\}$ a set of classes C = $\{c_1, c_2, \ldots\ldots, c_k\}$, the classification problem is to define a mapping $f: X \rightarrow C$, Where each x_i is assigned to one class.
- Naïve Bayes algorithm is a supervised learning algorithm, which is based on **Bayes theorem** and used for solving classification problems
- Probabilistic Approach to Learning. Instead of learning F: X → C, learn P(C|X).
- can design algorithms that learn functions with uncertain outcomes



Applications

- Face Recognition
- Weather Prediction
- Medical Diagnosis
- News Classification
- …

Example: Normal / Spam mail Classification

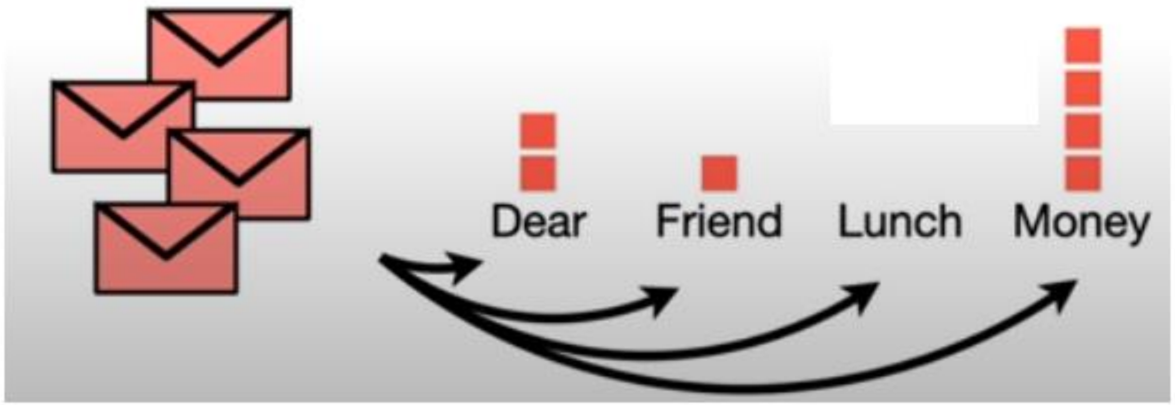


$$p(\textbf{Dear} \mid \text{Normal}) = \frac{8}{17} = 0.47$$

$$p(\textbf{Friend} \mid \text{Normal}) = \frac{5}{17} = 0.29$$

$$p(\textbf{Lunch} \mid \text{Normal}) = \frac{3}{17} = 0.18$$

$p(\textbf{Money} \mid N) = 0.06$



$p(\textbf{Dear} \mid S) = 0.29$

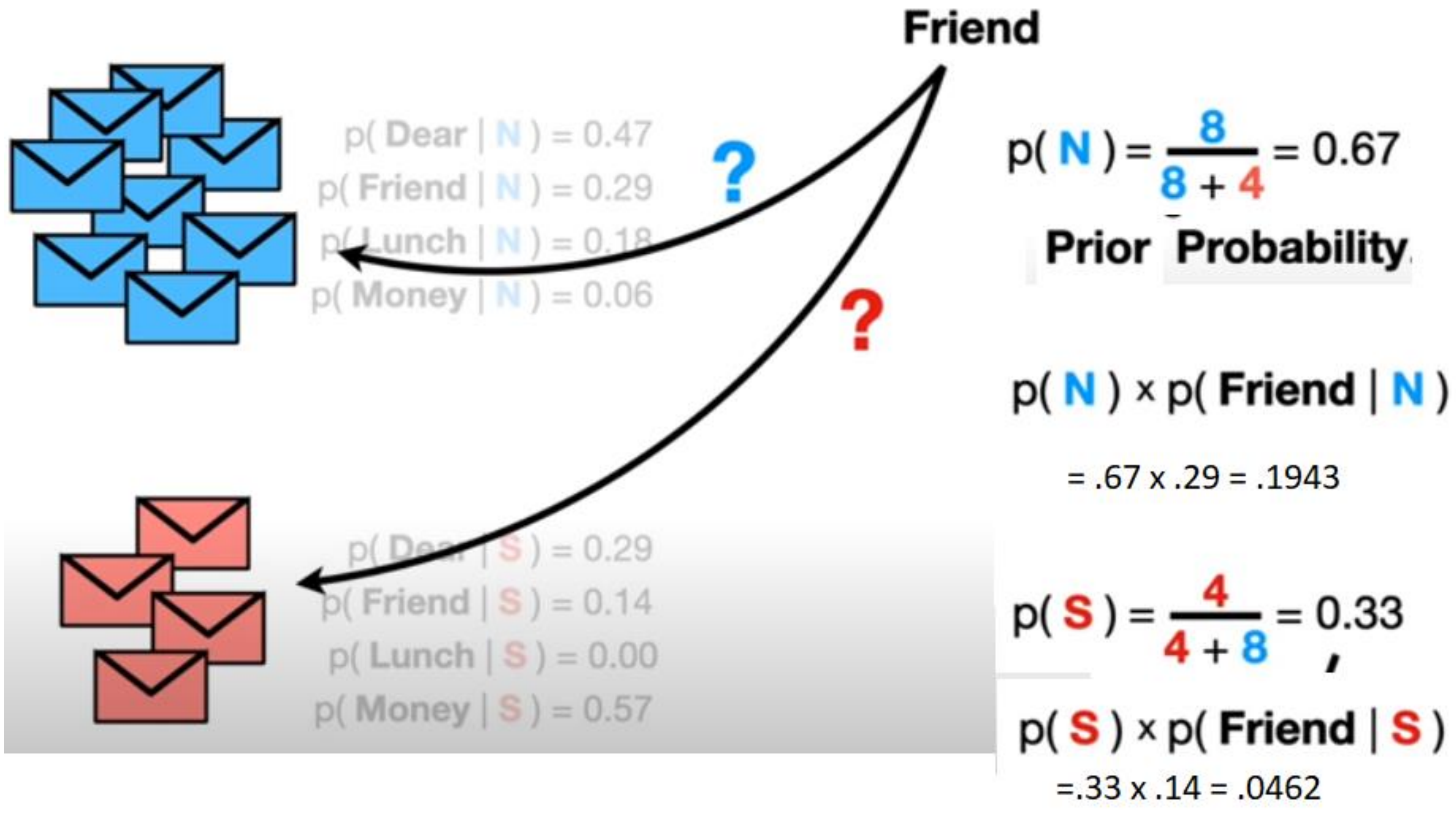
$p(\textbf{Friend} \mid S) = 0.14$

$p(\textbf{Lunch} \mid S) = 0.00$

$p(\textbf{Money} \mid S) = 0.57$

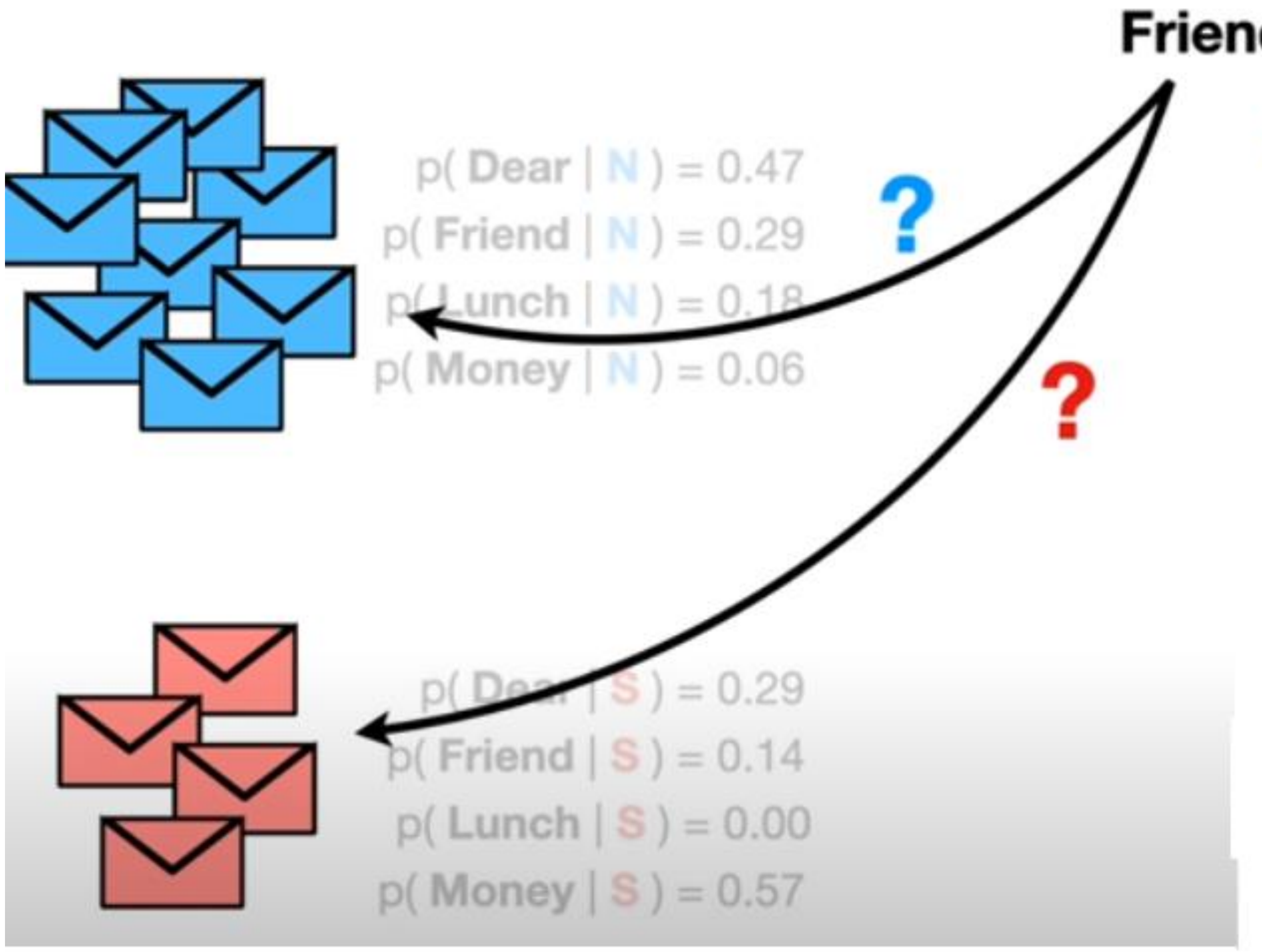
Slides taken from Josh Starmer, statquest

Example: Normal / Spam mail Classification



Slides taken from Josh Starmer, statquest

Example: Normal / Spam mail Classification



P(N|Friend)

p(N) x p(Friend | N)

= .67 x .29 = .1943

P(S|Friend)

p(S) x p(Friend | S)

=.33 x .14 = .0462

Slides taken from Josh Starmer, statquest

Naïve Bayesian Classifier

- Suppose, c is a class variable and $X = \{X_1, X_2, \ldots\ldots, X_n\}$ is a set of attributes, with instance of *c*.
- Naïve Bayesian classifier calculate this posterior probability using Bayes' theorem

$$\underbrace{P(c\,|\,x)}_{\text{Posterior Probability}} = \frac{\overbrace{P(x\,|\,c)}^{\text{Likelihood}}\,\overbrace{P(c)}^{\text{Class Prior Probability}}}{\underbrace{P(x)}_{\text{Predictor Prior Probability}}}$$

$$P(c\,|\,X) = P(x_1\,|\,c) \times P(x_2\,|\,c) \times \cdots \times P(x_n\,|\,c) \times P(c)$$

  - The probability $P(c|X)$ (also called class conditional probability) is therefore proportional to $P(X|c) \cdot P(c)$.
  - Thus, $P(c|X)$ can be taken as a measure of *c* given that *X*.

$$P(c/X) \approx P(X|c) \cdot P(c)$$

Weather data set

Outlook	Temp	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

Frequency Table		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3

https://www.saedsayad.com/naive_bayesian.htm

Likelihood Table

Frequency Table		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3

Likelihood Table		Play Golf		
		Yes	No	
Outlook	Sunny	3/9	2/5	5/14
	Overcast	4/9	0/5	4/14
	Rainy	2/9	3/5	5/14
		9/14	5/14	

$$P(x \mid c) = P(Sunny \mid Yes) = 3/9 = 0.33$$

$$P(x) = P(Sunny) = 5/14 = 0.36$$

$$P(c) = P(Yes) = 9/14 = 0.64$$

$$P(c \mid x) = P(Yes \mid Sunny) = 0.33 \times 0.64 \div 0.36 = 0.60$$

Frequency Table

		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3

		Play Golf	
		Yes	No
Humidity	High	3	4
	Normal	6	1

		Play Golf	
		Yes	No
Temp.	Hot	2	2
	Mild	4	2
	Cool	3	1

		Play Golf	
		Yes	No
Windy	False	6	2
	True	3	3

Likelihood Table

		Play Golf	
		Yes	No
Outlook	Sunny	3/9	2/5
	Overcast	4/9	0/5
	Rainy	2/9	3/5

		Play Golf	
		Yes	No
Humidity	High	3/9	4/5
	Normal	6/9	1/5

		Play Golf	
		Yes	No
Temp.	Hot	2/9	2/5
	Mild	4/9	2/5
	Cool	3/9	1/5

		Play Golf	
		Yes	No
Windy	False	6/9	2/5
	True	3/9	3/5

Prediction on test data

Outlook	Temp	Humidity	Windy	Play
Rainy	Cool	High	True	?

$$P(Yes \mid X) = P(Rainy \mid Yes) \times P(Cool \mid Yes) \times P(High \mid Yes) \times P(True \mid Yes) \times P(Yes)$$

$$P(Yes \mid X) = 2/9 \times 3/9 \times 3/9 \times 3/9 \times 9/14 = 0.00529$$

$$P(No \mid X) = P(Rainy \mid No) \times P(Cool \mid No) \times P(High \mid No) \times P(True \mid No) \times P(No)$$

$$P(No \mid X) = 3/5 \times 1/5 \times 4/5 \times 3/5 \times 5/14 = 0.02057$$

Naïve Bayesian Classifier Algorithm

Input: Given a set of *k* mutually exclusive and exhaustive classes $C = \{c_1, c_2, \ldots\ldots, c_k\}$, which have prior probabilities $P(C_1), P(C_2), \ldots\ldots P(C_k)$.

There are *n*-attribute set $A = \{A_1, A_2, \ldots\ldots, A_n\}$, which for a given instance have values $A_1 = a_1$, $A_2 = a_2$,….., $A_n = a_n$

Step: For each $c_i \in C$, calculate the class condition probabilities, $i = 1,2,\ldots\ldots,k$

$$p_i = P(C_i) \times \prod_{j=1}^{n} P(A_j = a_j | C_i)$$

$$p_x = \max\{p_1, p_2, \ldots\ldots, p_k\}$$

Output: C_x is the classification

Naïve Bayesian Classifier Pros and Cons

Pros

- simple and easy to implement
- doesn't require as much training data
- handles both continuous and discrete data
- fast and can be used to make real-time predictions
- not sensitive to irrelevant features

Cons

- assumption of independent predictors
- zero frequency problem