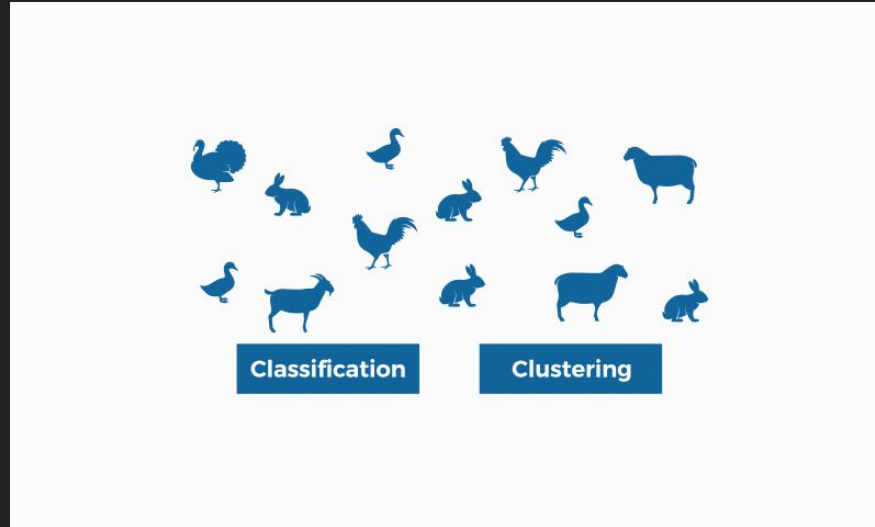


Classification vs. Clustering



Classification

- Predetermined classes
- User defined
-

Clustering

- Unsupervised classes
- Machine defined
-

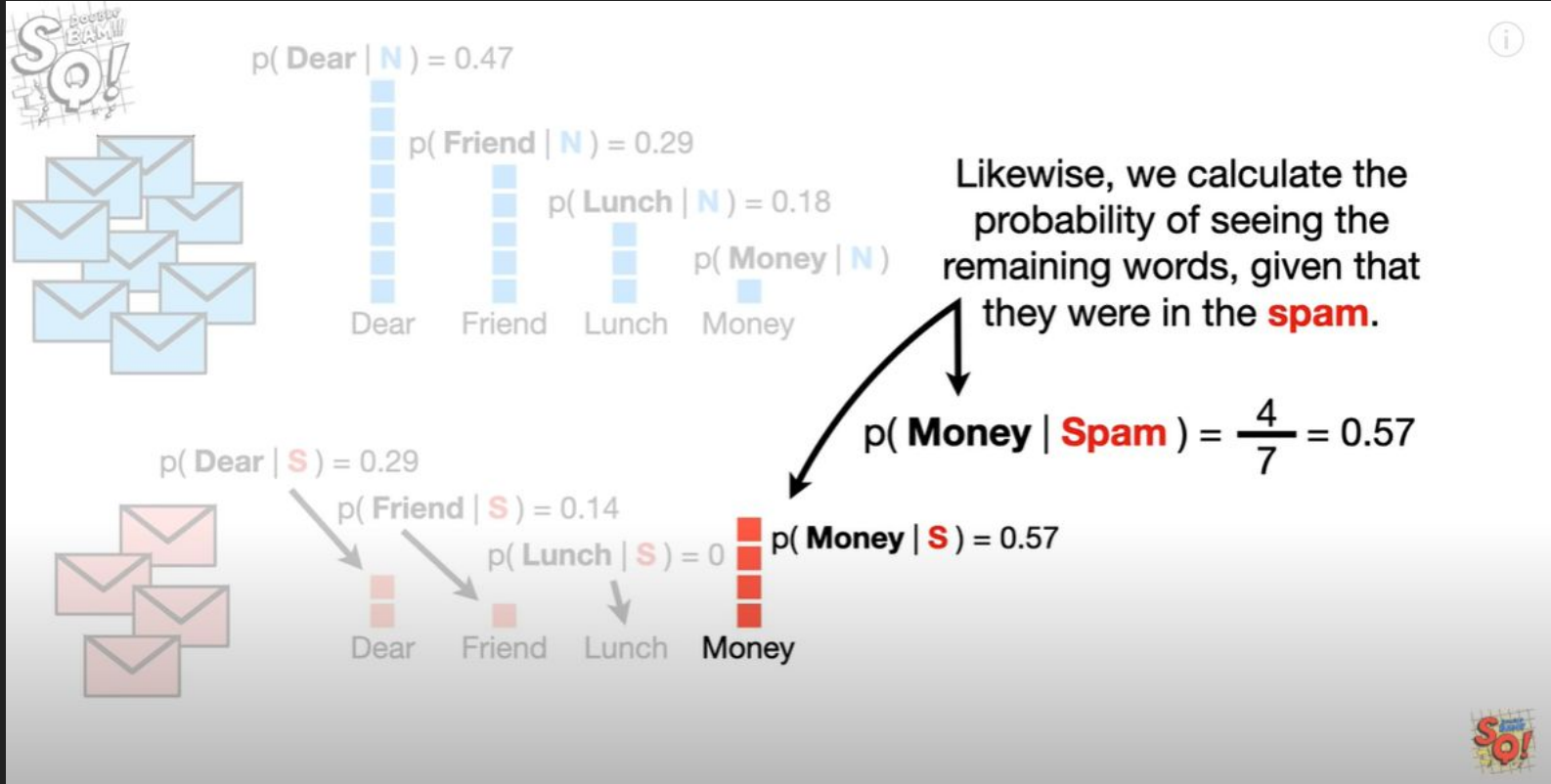
Naive Bayes Classifiers

- Classification Technique
- Based on Bayes' Theorem
- Naive Bayes classifiers assume that the presence of a particular feature in a class is unrelated to the presence of any other feature.



- Red
- Round
- 3 Inches in Diameter

Emails: Naive Bayes Theorem



Emails: Naive Bayes Theorem



$$\begin{aligned}p(\text{Dear} \mid \text{N}) &= 0.47 \\p(\text{Friend} \mid \text{N}) &= 0.29 \\p(\text{Lunch} \mid \text{N}) &= 0.18 \\p(\text{Money} \mid \text{N}) &= 0.06\end{aligned}$$

Dear Friend

?

?

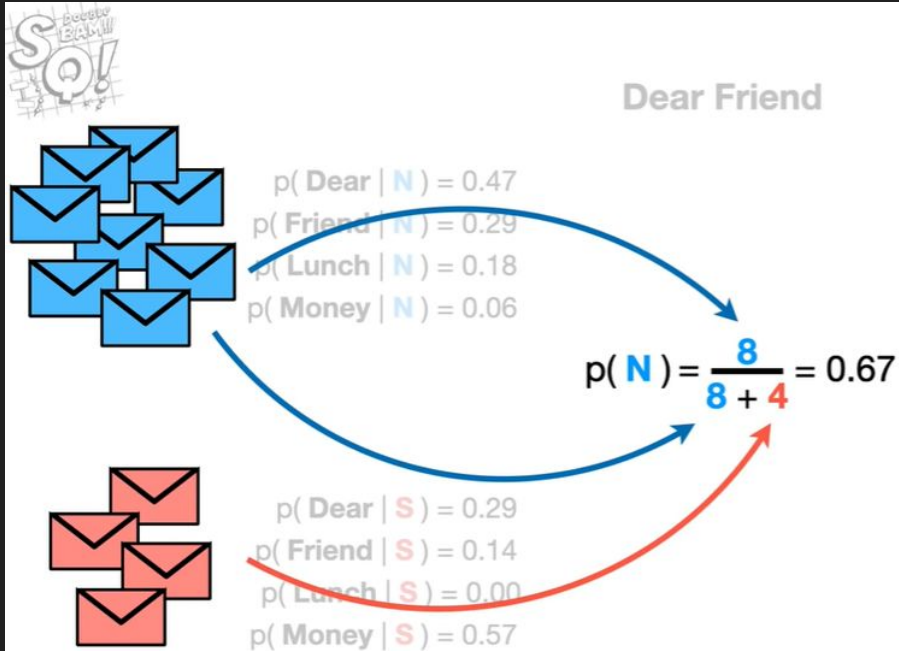


$$\begin{aligned}p(\text{Dear} \mid \text{S}) &= 0.29 \\p(\text{Friend} \mid \text{S}) &= 0.14 \\p(\text{Lunch} \mid \text{S}) &= 0.00 \\p(\text{Money} \mid \text{S}) &= 0.57\end{aligned}$$

And we want to decide if is a **normal message** or **spam**.



Emails: Naive Bayes Theorem



For example, since **8** of the **12** messages are **normal messages**, our initial guess will be **0.67**.



Emails: Naive Bayes Theorem

Normal Messages (N)

- $p(N) = 0.67$
- $p(\text{Dear} | N) = 0.47$
- $p(\text{Friend} | N) = 0.29$
- $p(\text{Lunch} | N) = 0.18$
- $p(\text{Money} | N) = 0.06$

Spam Messages (S)

- $p(\text{Dear} | S) = 0.29$
- $p(\text{Friend} | S) = 0.14$
- $p(\text{Lunch} | S) = 0.00$
- $p(\text{Money} | S) = 0.57$

Sample Email: Dear **Friend**

...and the probability that the word **Friend** occurs in a **normal message**.

Joint Probability Formula:

$$p(N) \times p(\text{Dear} | N) \times p(\text{Friend} | N)$$

Emails: Naive Bayes Theorem



$$p(\mathbf{N} \mid \text{Dear Friend}) \propto 0.09$$

Dear Friend



$$p(\mathbf{N}) = 0.67$$

$$\begin{aligned} p(\text{Dear} \mid \mathbf{N}) &= 0.47 \\ p(\text{Friend} \mid \mathbf{N}) &= 0.29 \\ p(\text{Lunch} \mid \mathbf{N}) &= 0.18 \\ p(\text{Money} \mid \mathbf{N}) &= 0.06 \end{aligned}$$

So let's put that on top of the **normal messages** so we don't forget.

$$0.67 \times 0.47 \times 0.29 = 0.09 \propto p(\mathbf{N} \mid \text{Dear Friend})$$



$$\begin{aligned} p(\text{Dear} \mid \mathbf{S}) &= 0.29 \\ p(\text{Friend} \mid \mathbf{S}) &= 0.14 \\ p(\text{Lunch} \mid \mathbf{S}) &= 0.00 \\ p(\text{Money} \mid \mathbf{S}) &= 0.57 \end{aligned}$$



Emails: Naive Bayes Theorem



$$p(\text{N} \mid \text{Dear Friend}) \propto 0.09$$

Dear Friend



$$p(\text{N}) = 0.67$$

$$\begin{aligned} p(\text{Dear} \mid \text{N}) &= 0.47 \\ p(\text{Friend} \mid \text{N}) &= 0.29 \\ p(\text{Lunch} \mid \text{N}) &= 0.18 \\ p(\text{Money} \mid \text{N}) &= 0.06 \end{aligned}$$

However, technically, it is
proportional to the probability
that the message is **spam**
given that it says **Dear Friend**.

$$0.33 \times 0.29 \times 0.14 = 0.01 \propto p(\text{S} \mid \text{Dear Friend})$$



$$p(\text{S}) = 0.33$$

$$\begin{aligned} p(\text{Dear} \mid \text{S}) &= 0.29 \\ p(\text{Friend} \mid \text{S}) &= 0.14 \\ p(\text{Lunch} \mid \text{S}) &= 0.00 \\ p(\text{Money} \mid \text{S}) &= 0.57 \end{aligned}$$

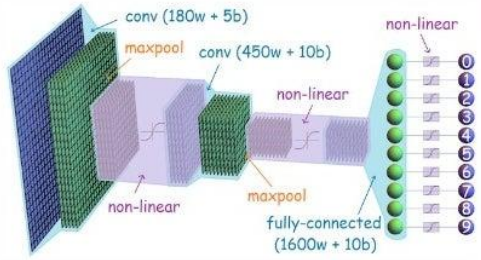


Why so . . . Naive?

- Treats all word orders the same.
- The score for the phrase “Friend Dear” and “Dear Friend” - the exact same.
- Like picking at random from a bag of words.
- But performs very well when ignoring spam
- High Bias
- Low Variance

WHO WOULD WIN?

AN INCREDIBLY COMPLEX MULTI-LAYER CONVOLUTIONAL NEURAL NETWORK



The diagram illustrates a complex multi-layer convolutional neural network (CNN) architecture. It starts with an input layer (blue grid) followed by a convolutional layer (conv) with parameters (180w + 5b), a maxpooling layer (maxpool), another convolutional layer (conv) with parameters (450w + 10b), a non-linear activation layer (non-linear), a third maxpooling layer (maxpool), a fourth non-linear activation layer (non-linear), and finally a fully-connected layer (fully-connected) with parameters (1600w + 10b). The output is a vector of 10 nodes, labeled 0 through 9, representing the final classification results.

ONE NAIVE BOI



A meme image featuring a black and white portrait of a man wearing a flat cap, with the text "ONE NAIVE BOI" overlaid in a bold, white, sans-serif font.