Naive bayes

martes, 2 de noviembre de 2021 3:06 p. m.

Histograma — Probabilidades

↳ Ejemplo correo spam.

p(**Dear** | **N**) = 0.47
p(**Friend** | **N**) = 0.29
p(**Lunch** | **N**) = 0.18
p(**Money** | **N**) = 0.06

Now, because these histograms are taking up a lot of space, let's get rid of them, but keep the probabilities.

p(**Dear** | **S**) = 0.29
p(**Friend** | **S**) = 0.14
p(**Lunch** | **S**) = 0.00
p(**Money** | **S**) = 0.57

} Likelihoods / Probabilities

"Dear friend"

$p(N) = 0.67$ $\qquad p(N) = \frac{8}{8+4} = 0.67$

↳ prior probability

Normal classage

0.09

↳ Normal message

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

$$p(S) \times p(\text{Dear} \mid S) \times p(\text{Friend} \mid S) \rightarrow 0.01$$

2

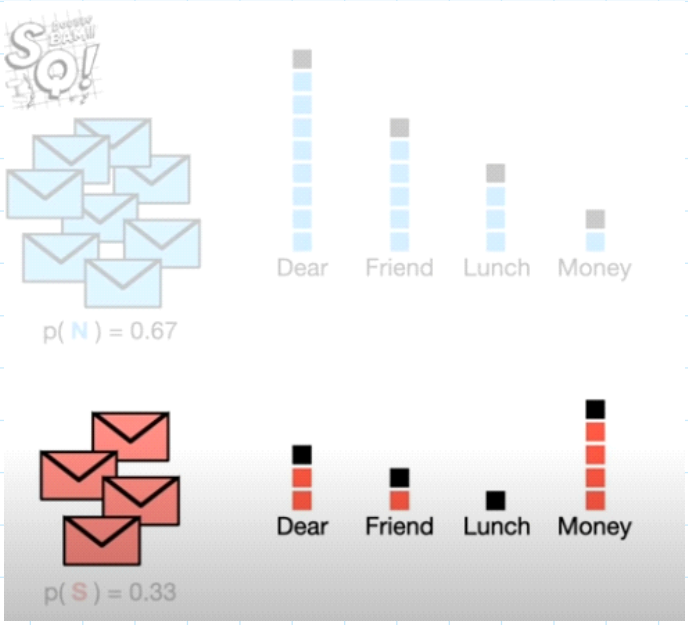
Lunch **Money Money Money Money**

$$p(N) \times p(\text{Lunch} \mid N) \times p(\text{Money} \mid N)^4 \simeq 0.00002$$

$p(N) \times p(\text{Lunch} \mid N) \times p(\text{Money} \mid N)^4 \approx 0.00002$

$p(S) \times p(\text{Lunch} \mid S) \times p(\text{Money} \mid S)^4 = 0 \quad = 0$

para evitar eso → 0 ⇒ ¿cuánto?



útil //

$p(N) \times p(\text{Lunch} \mid N) \times p(\text{Money} \mid N)^4 = 0.00001$

$p(S) \times p(\text{Lunch} \mid S) \times p(\text{Money} \mid S)^4 = 0.00122$

↳ ¡SPAM!

Naive Bayes.

The thing that makes **Naive Bayes** so ***naive*** is that it treats all word orders the same.

Score for **Dear Friend** =

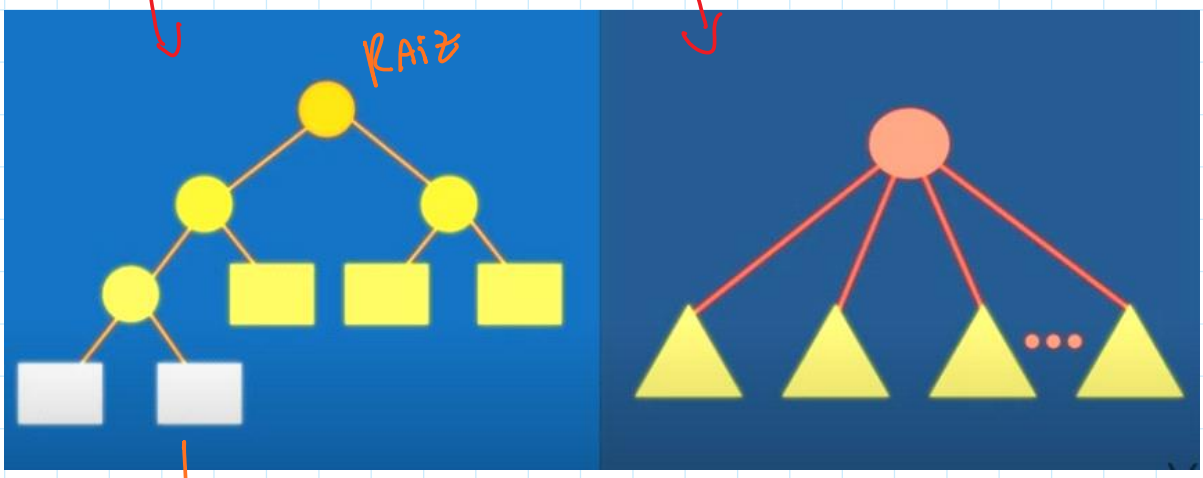
$p(N) \times p(\text{Dear} \mid N) \times p(\text{Friend} \mid N) = 0.08$

Score for **Friend Dear** =

$p(N) \times p(\text{Friend} \mid N) \times p(\text{Dear} \mid N) = 0.08$

Ignora las reglas de lenguaje

Decision Tree y Random Forest



Preguntas, Arbol decisión

Dataset / Etiquetas y rasgos

				E
1	0	1	0	1
1	1	0	0	0
1	1	1	0	1

Incompletos

Gini Impurity,

Menor coeficiente → Nodo raiz

Arbol de decisión

Datos al Azar

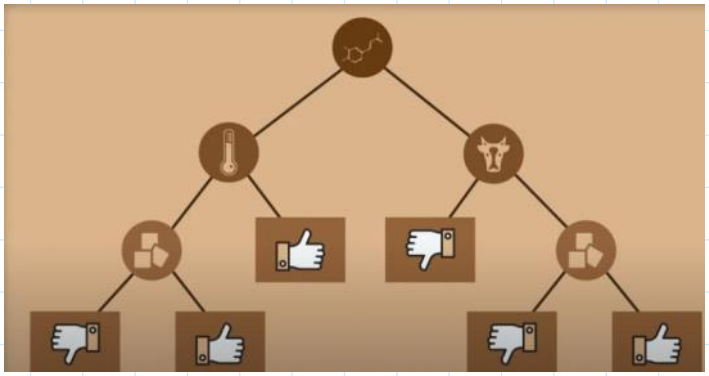
Arbol de decisión ~ veces

Varios tambien

Dato clasificado ~ veces

100

↳ Nodo raíz



↳ Clasificar Datos
↳ Numericos y Booleanos
↳ Un clasificador

↳ Dependiendo de los n árboles
↳ "Democracia"
↳ Arboles mejor criterio
↳ Datos de prueba
↳ Se pueden combinar los subconjuntos
↳ Varios clasificadores
↳ Difícil escoger