

Supongamos q' tenemos un conjunto de palabras.

(2)

$W = \{w_1, \dots, w_n\}$; ~~evento~~ X_i evento "el mail contiene la palabra w_i ."

y supongamos q' podemos estimar $P(X_i/S)$ y $P(X_i/\bar{S})$.

Notar q' $X_i = \{0, 1\}$

El modelo Naive Bayes hace una asunción extrema: q' los sucesos son independientes, es decir:

$$P(X_1=x_1, \dots, X_n=x_n/S) = P(X_1=x_1/S) \dots P(X_n=x_n/S) = P(X=x/S)$$

O sea, por ejemplo: la probabilidad de que un mail con spam contenga la palabra "viagra" o la palabra "rolex" son sucesos independientes.