An Example of Text Classification with Naïve Bayes

Dell Zhang
11/11/2006

Two Classes: ham and spam

Training Data

ham d_1: "good."
ham d_2: "very good."
spam d_3: "bad."
spam d_4: "very bad."
spam d_5: "very bad, very bad."

Test Data

? d_6: "good? bad! very bad!"

// Naïve Bayes Learning

// Prior Probability

$P(\text{ham}) \quad = N_{\text{ham}} \ / \ (N_{\text{ham}}+N_{\text{spam}}) = 2 \ / \ (2+3) = 0.40$
$P(\text{spam}) \quad = N_{\text{spam}} \ / \ (N_{\text{ham}}+N_{\text{spam}}) = 3 \ / \ (2+3) = 0.60$

// Constructing the Unigram Language Model
// Add-One Smoothing

$P(\text{very}|\text{ham})$
$= (T_{\text{very}|\text{ham}}+1) \ / \ ((T_{\text{very}|\text{ham}}+1) + (T_{\text{good}|\text{ham}}+1) + (T_{\text{bad}|\text{ham}}+1))$
$= (1+1) \ / \ ((1+1) + (2+1) + (0+1))$
$= 2 \ / \ (2 + 3 + 1)$
$= 0.33$

$P(\text{good}|\text{ham})$
$= (T_{\text{good}|\text{ham}}+1) \ / \ ((T_{\text{very}|\text{ham}}+1) + (T_{\text{good}|\text{ham}}+1) + (T_{\text{bad}|\text{ham}}+1))$
$= (2+1) \ / \ ((1+1) + (2+1) + (0+1))$
$= 3 \ / \ (2 + 3 + 1)$
$= 0.50$

$P(\text{bad}|\text{ham})$
$= (T_{\text{bad}|\text{ham}}+1) \ / \ ((T_{\text{very}|\text{ham}}+1) + (T_{\text{good}|\text{ham}}+1) + (T_{\text{bad}|\text{ham}}+1))$
$= (0+1) \ / \ ((1+1) + (2+1) + (0+1))$
$= 1 \ / \ (2 + 3 + 1)$
$= 0.17$

$P(\text{very}|\text{spam})$
$= (T_{\text{very}|\text{spam}}+1) \ / \ ((T_{\text{very}|\text{spam}}+1) + (T_{\text{good}|\text{spam}}+1) + (T_{\text{bad}|\text{spam}}+1))$
$= (3+1) \ / \ ((3+1) + (0+1) + (4+1))$
$= 4 \ / \ (4 + 1 + 5)$
$= 0.40$

$P(\text{good}|\text{spam})$
$= (T_{\text{good}|\text{spam}}+1) \ / \ ((T_{\text{very}|\text{spam}}+1) + (T_{\text{good}|\text{spam}}+1) + (T_{\text{bad}|\text{spam}}+1))$
$= (0+1) \ / \ ((3+1) + (0+1) + (4+1))$
$= 1 \ / \ (4 + 1 + 5)$
$= 0.10$

$P(\text{bad}|\text{spam})$
$= (T_{\text{bad}|\text{spam}}+1) \ / \ ((T_{\text{very}|\text{spam}}+1) + (T_{\text{good}|\text{spam}}+1) + (T_{\text{bad}|\text{spam}}+1))$

$$= (4+1) / ((3+1) + (0+1) + (4+1))$$
$$= 5 / (4 + 1 + 5)$$
$$= 0.50$$

// Naïve Bayes Classification

// Likelihood
// Applying the Unigram Language Model

$P(d_6|\text{ham})$
$= \mathrm{P}(\text{good}|\text{ham})\times\mathrm{P}(\text{bad}|\text{ham})\times\mathrm{P}(\text{very}|\text{ham})\times\mathrm{P}(\text{bad}|\text{ham})$
$= 0.50\times0.17\times0.33\times0.17$
$= 0.0048$

$P(d_6|\text{spam})$
$= \mathrm{P}(\text{good}|\text{spam})\times\mathrm{P}(\text{bad}|\text{spam})\times\mathrm{P}(\text{very}|\text{spam})\times\mathrm{P}(\text{bad}|\text{spam})$
$= 0.10\times0.50\times0.40\times0.50$
$= 0.010$

// Posterior Probability

$P(\text{ham}|\ d_6)$
$= \mathrm{P}(d_6|\text{ham})\times\mathrm{P}(\text{ham})/\mathrm{P}(d_6)$ // Bayes' Rule
$= 0.0048\times0.40/\mathrm{P}(d_6)$
$= 0.0019/\mathrm{P}(d_6)$

$P(\text{spam}|\ d_6)$
$= \mathrm{P}(d_6|\text{spam})\times\mathrm{P}(\text{spam})/\mathrm{P}(d_6)$ // Bayes' Rule
$= 0.010\times0.60/\mathrm{P}(d_6)$
$= 0.0060/\mathrm{P}(d_6)$

// Classification

Since $P(\text{ham}|\ d_6) < P(\text{spam}|\ d_6)$,
d_6 should be classified into the spam class.

P.S.

Note that without Add-One (Laplace) smoothing,
$P(\text{good}|\text{spam})$ would be 0 and consequently $\mathrm{P}(\text{spam}|\ d_6)$ would be 0.