Naïve Bayes Classification

- We could try to find the maximum probability given some conditions

$$p(positive|word1, word2, word3, \dots)$$

- But we do not have enough data

- So start to break this down using Bayes' Law, using c to represent the classification and $\mathbf{x}$ the conditions:

$$p(c|\mathbf{x}) = \frac{p(c)p(\mathbf{x}|c)}{p(\mathbf{x})}$$

- We only wish to maximise this, so can drop the $p(\mathbf{x})$

Naïve Bayes Classification

- Can chain conditional probabilities but it gets ugly:

$$p(x_1, x_2, x_3, \ldots | c) = p(x_1|c)p(x_2, x_3, \ldots | c, x_1)$$

$$= p(x_1|c)p(x_2|c, x_1)p(x_3, \ldots | c, x_1, x_2) \ldots$$

- In Naïve Bayes classification, we assume the features are conditionally independent given the classification

- That is:

$$p(x_i|c, x_{i-1}) = p(x_i|c)$$

- Hence:

$$p(\mathbf{x}|c) = p(x_1|c)p(x_2|c) \ldots p(x_N|c)$$