Naive Bayes

- Maximum Likelihood Estimation (MLE)
  - Consider a sequence of coin flips, for example Coin
  - prior → HHTTTTTTHTTTTTHTTHTT (5 times H, 15 times T)
  - Which Pr(H) and Pr(T) are the most likely?
  - Looks like Pr(H) = ¼ and Pr(T) = ¾ ...

$Pr(H) = Pr(T) = 1/2$

$|\#\text{flips}| = 20$

$|H| = 5, \ |T| = 15$

$P(T) = \frac{|T|}{|F|} = 15/20 = 3/4$

$P(H) = \frac{|H|}{|F|} = 5/20 = 1/4$

- Conditional probabilities
  - Let A and B be events in a probability space Ω
  - Denote by Pr(A | B) the probability of A ∩ B in the space B

→ **(1)** Pr(A | B) := Pr(A ∩ B) / Pr (B)

→ **(2)** Pr(A | B) · Pr(B) = Pr (B | A) · Pr(A)

$\Omega = \{1, 2, 3, 4, 5, 6\}$

$A = \{2, 4, 6\} \Rightarrow P(A) = \frac{|A|}{|\Omega|} = \frac{3}{6} = \frac{1}{2}$

$B = \{1, 2, 3\} \Rightarrow P(B) = \frac{|B|}{|\Omega|} = 3/6 = 1/2$

$P(A|B) = \frac{|A \cap B|}{|B|} = \frac{1}{3}$ ✓

$Pr(A|B) = Pr(B|A) \cdot P(A) / P(B)$

- Probabilistic assumptions
  - Underlying probability distributions:

    A distribution p_c over the classes ... where $\Sigma_c\, p_c = 1$

    For each c, a distr. p_{wc} over the words ... where $\Sigma_w\, p_{wc} = 1$
  - Naïve Bayes assumes the following process for generating a document D with m words $W_1 \ldots W_m$ and class label C:

    Pick C=c with prob. p_c , then pick each word W_i=w with probability p_{wc} , independent of the other words

- Learning phase
  - For a **training set** T of objects, let:

    T_c = the set of documents from class c

    n_{wc} = #occurrences of word w in documents from T_c

    n_c = #occurrences of all words in documents from T_c
  - We compute the following probabilities or likelihoods:

→ $\mathbf{p_c} \coloneqq |\mathbf{T_c}| \;/\; |\mathbf{T}|$ global likelihood of a class

$\mathbf{p_{wc}} \coloneqq \mathbf{n_{wc}} \;/\; \mathbf{n_c}$ likelihood of a word for a class

$T = 90$

→ $T_{Horror} = 10$

→ $T_{Doc} = 25$

→ $T_{Comedy} = 15$

$P_{\text{Jupyter, Doc}} = \frac{20}{80} = 1/4 = 0.25$

- Learning phase, example
  - Consider Example 2 (artificial documents)

aba	A
baabaaa	A
bbaabbab	B
abbaa	A
abbb	B
bbbaab	B

$T_A = 3$, $T_B = 3$, $T = 6$

$P_A = P(T_A) = 3/6 = 1/2$, $\boxed{P_B = 1/2}$

$n_{aA} = 10$, $n_{bA} = 5$, $n_A = 15$

$P_{aA} = \frac{10}{15} = \boxed{2/3}$, $P_{bA} = 5/15 = \boxed{1/3}$

$n_{aB} = 6$, $n_{bB} = 12$, $n_B = 18$

$P_{aB} = 6/18 = 1/3$, $P_{bB} = 12/18 = 2/3$

$16/8 < 24/8$

2 3

0.5

- Prediction
  - For a given document **d** we want to compute the probability of each class **c**, given document d:

    Pr(C=c | D=d)

  - Using Bayes Theorem, we have:

    Pr(C=c | D=d) = Pr(D=d | C=c) · Pr(C=c) / ~~Pr(D=d)~~

  - Using our (naive) probabilistic assumptions, we have:

    $\Pr(D=d \mid C=c) = \Pr(W_1=w_1 \cap \ldots \cap W_m=w_m \mid C=c)$

    $= \Pi_{i=1,\ldots,m} \Pr(W_i=w_i \mid C=c)$

    A

    a, b

- Prediction … continued
  - We thus obtain that

    $\Pr(C=c \mid D=d)$

    $= \Pi_{i=1,\ldots,m} \Pr(W_i=w_i \mid C=c) \cdot \Pr(C=c) / \Pr(D=d)$

    $= \mathbf{\Pi_{i=1,\ldots,m}\ p_{w_i c} \cdot p_c} / \Pr(D=d)$

  - Note 1: for $\Pi_{i=1,\ldots,m}\ p_{w_i c}$ just take the p_{wc} for all words w in the document and multiply them (if a word w occurs multiple times, also take the factor p_{wc} multiple times)
  - Note 2: the Pr(D=d) is the same for all classes c

- Prediction, example
  - Consider Example 2 (artificial documents)

aba	A
baabaaa	A
bbaabbab	B
abbaa	A
abbb	B
bbbaab	B

  - Let us predict the class for aab … A or B ?

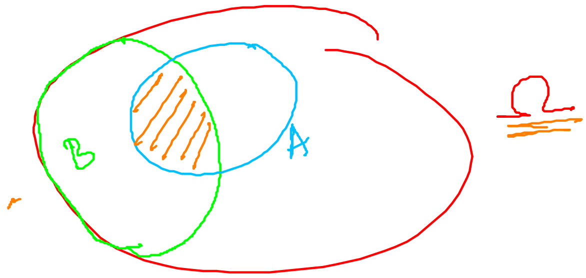
$P(A \mid aab) = P_{aA} \cdot P_{aA} \cdot P_{bA} \cdot P_A$

$= 2/3 \cdot 2/3 \cdot 1/3 \cdot 1/2 = 2/27$

$P(B \mid aab) = P_{aB} \cdot P_{aB} \cdot P_{bB} \cdot P_B$

$1/3 \cdot 1/3 \cdot 2/3 \cdot 1/2 = 1/27$

$P(A/aab) > P(B \mid aab) \rightarrow$ Predict A



$$P(A|B) = \frac{|A \cap B|}{|B|}$$

$$= \frac{|A \cap B| \,/\, |\Omega|}{|B| \,/\, |\Omega|} = \frac{Pr(A \cap B)}{Pr(B)}$$

$$P(\text{Zombie}, \text{Horror}) = \frac{|\text{Zombie}|}{|\text{class, words}|} = \frac{5}{20} = 1/4 = 0.25$$