Naive Bayes

- **Classifies** vectors of discrete-valued features, features are ^{ASSUMED} conditionally independent w.r.t label

- $P(x|y=c_i, \theta) = \prod_{j}^{D} p(x_j|y=c, \theta_{jc})$ • **RESILIENT TO OVERFITTING** and curse of dimensionality

- Class conditional densities / <u>LIKELIHOODS</u> can be real (<u>Gaussian</u>), binary (<u>Bernoulli</u>), categorical (<u>Multinoulli</u>)

- base case, MLE, counts proportion

- Factored prior to avoid crash when new words

- **FEATURE SELECTION w/ MUTUAL INFORMATION:** $I(x_i, y) = \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x_i)p(y)}$

- **IF FEATURES ARE C.I.** → Naive Bayes returns MAP

- $P(c_n|x_1 \dots x_w) \longrightarrow P(c_n|x) = \frac{P(c_n)P(x|c_n)}{P(x)}$ $\xrightarrow{c_i}$ $\propto P(c_n) \prod^{N} p(x_i|c_n)$ $\xrightarrow{\text{DECISION}}$ $\underset{k\{1 \dots n\}}{ARGMAX}[\dots]$

- Can regularize with Laplace smoothing or Lidstone smoothing

- **MULTINOMIAL NBC** → **LINEAR CLASSIFIER** in logspace

- Generative / discriminative pair w/ logistic regression

- Laplace smoothing → add counts to $n, 0$ to avoid it freaking up when sees stuff not in training set