**Summary of Naïve Bayes**

* Hypotheses: H1 through Hn.
* Features (data): F1 through Fm.
* 
* P(*Hi*) for i = 1 to n.
* P(*Fj* | *Hi*) for j = 1 to m and i = 1 to n.

**Spam example 1**

* Suppose I know 80% of my email is spam.
* I have three features, "luxury," "brands," and "save."
* I know:
  + P(luxury | spam) = 0.4 P(luxury | ~spam) = 0.01
  + P(brands | spam) = 0.3 P(brands | ~spam) = 0.2
  + P(save | spam) = 0.4 P(save | ~spam) = 0.1
* Suppose a new, incoming email contains "luxury" and "save" but not "brands." Should it be classified as spam or ~spam?

**Spam example 2**

Suppose I have 20 emails that have been already classified into spam (15 emails) and non-spam (5 emails). Suppose I only care about the presence or absence of the words **luxury**, **brands**, and **save**.  
  
Suppose 6 of the spam emails contain "luxury," 3 of the spam emails contain "brands," and 7 of the spam emails contain "save."  
  
Suppose 1 of the non-spam emails contains "luxury," 2 of the non-spam emails contain "brands," and 2 of the non-spam emails contain "save."  
  
Suppose a new email arrives that contains the words "luxury" and "save" but not "brands." Should this be classified as spam or not spam?