Assignment

## Spam vs. Ham with Naive Bayes

**Training corpus**

| Doc | Words | Label |
| --- | --- | --- |
| 1 | special offer | Spam |
| 2 | family photo | Ham |
| 3 | special meeting | Ham |
| 4 | offer inside | Spam |

**New document**: “special photo”

### 1. Prior probabilities

Total docs = 4  
Spam docs = 2 → P(Spam) = 2/4 = 0.5  
Ham docs = 2 → P(Ham) = 2/4 = 0.5

### 2. Likelihoods (Laplace smoothing, α = 1)

**Spam vocabulary size V\_S = 4** {special, offer, offer, inside}  
**Ham vocabulary size V\_H = 4** {family, photo, special, meeting}

| Term | Class | Count in class | P(word |
| --- | --- | --- | --- |
| special | Spam | 1 | (1+1)/(4+4) = 2/8 = 0.25 |
| photo | Spam | 0 | (0+1)/(4+4) = 1/8 = 0.125 |
| special | Ham | 1 | (1+1)/(4+4) = 2/8 = 0.25 |
| photo | Ham | 1 | (1+1)/(4+4) = 2/8 = 0.25 |

### 3. Posterior calculation (ignoring shared denominator)

**Spam score**  
P(Spam) · P(special|Spam) · P(photo|Spam)  
= 0.5 × 0.25 × 0.125 = 0.015625

**Ham score**  
P(Ham) · P(special|Ham) · P(photo|Ham)  
= 0.5 × 0.25 × 0.25 = 0.03125

### 4. Decision

Ham score > Spam score ⇒ classify **“special photo” as Ham**.