Patern Recognition Principles Week6 Assignment

#### Training Data Summary

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

New document to classify: **“special photo”**

### 1. Prior Probabilities

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

### 2. Likelihoods (word-probabilities given class)

We use **word-frequency** counts inside each class (add-one smoothing omitted here to keep numbers simple).

#### Spam corpus words: {special, offer, offer, inside} → 4 tokens

* P(“special” | Spam) = 1/4 = **0.25**
* P(“photo” | Spam) = 0/4 = **0**

#### Ham corpus words: {family, photo, special, meeting} → 4 tokens

* P(“special” | Ham) = 1/4 = **0.25**
* P(“photo” | Ham) = 1/4 = **0.25**

### 3. Classification of “special photo”

Under the conditional-independence assumption:

#### Spam score (proportional to posterior)

#### Ham score (proportional to posterior)

Since 0.03125 > 0, the document **“special photo” is classified as Ham**.