**Solutions to Homework 2. Decision Theory and Bayesian Inference**

An organization uses a spam filtering software to block email messages that may potentially be spam messages. The spam filter can be set to one of two security modes: High-Security-Mode () or Low-Security-Mode ().

Extensive evaluation using a benchmark corpus consisting exclusively of *spam* messages yields the following performance statistics for the spam filter:

* 96% of the *spam* messages are blocked in the High-Security-Mode.
* 90% of the *spam* messages are blocked in the Low-Security-Mode.

Extensive evaluation using a benchmark corpus consisting exclusively of *non-spam* (legitimate) messages yields the following performance statistics for the spam filter:

* 10% of the *non-spam* messages are blocked in the High-Security-Mode
* 4% of the *non-spam* messages are blocked in the Low-Security-Mode

The organization estimates that 80% of the messages that it receives are *spam* messages.

Define: : Event that a message is a spam. : Event that a message is blocked.

Given:

|  |  |  |  |
| --- | --- | --- | --- |
| **High-Security-Mode** | | **Low-Security-Mode** | |
|  |  |  |  |
|  |  |  |  |

1. Let denote the conditional probability that a message that is not blocked by the spam filter operating in the High-Security-Mode is actually a spam message. Estimate .
2. Let denote the conditional probability that a message that is blocked by the spam filter operating in the High-Security-Mode is actually not a spam message. Estimate .
3. Let denote the conditional probability that a message that is not blocked by the spam filter operating in the Low-Security-Mode is actually a spam message. Estimate .
4. Let denote the conditional probability that a message that is blocked by the spam filter operating in the Low-Security-Mode is actually not a spam message. Estimate .

1. If the cost of not blocking a *spam* message is $1 and the cost of blocking a *non-spam* message is $10, should the organization operate the spam filter in the High-Security-Mode? Why?

Substituting the given probabilities, we have:

The organization should NOT operate the spam filter in the High-Security-Mode since the expected cost is higher.

1. Recall that the cost of not blocking a *spam* message is $1. At least how high should the cost of blocking a *non-spam* message be for a risk-neutral rational decision maker to prefer operating the spam filter in the Low-Security-Mode?

Let the cost of blocking a *non-spam* message = $ .

A risk-neutral rational decision maker to prefer operating the spam filter in the Low-Security-Mode when its expected cost is lower.

That is when:

That is when

1. Let be the amortized cost per message of operating the spam filter. Write a short memo the CIO of the organization explaining at most how high could be for the organization to use the spam filter with the specified performance?

We operate the spam filter in the Low-Security-Mode since the expected cost per message is lower ($0.16) than when operating it in the High-Security-Mode. Without a spam filter in place we won’t be blocking any messages. Under our current assumptions, the resulting expected cost per message is $ since 80% of the messages are spam and the cost for not blocking a spam is estimated to be a dollar. Thus, the spam filter saves us per message. We should be willing to invest in the spam filter as long as its amortized cost per message is less than $0.64