* [Overview](#gjdgxs)
* [Introduction](#30j0zll)
* [Reading](#1fob9te)

# CSE 450 - Machine Learning & Data Mining

## Preparation Reading 05: Bayes' Theorem and the Base Rate Fallacy

## Overview

In this reading you'll learn about Bayes' Theorem and how it applies to Machine Learning.

Estimated Reading Time

Plan on around 60 to 90 minutes for this preparation reading, which contains a mix of online and textbook reading.

## Introduction

Consider the following question:

If a test to detect a disease whose prevalence is 1/1,000 has a false positive rate of 5 per cent, what is the chance that a person found to have a positive result actually has the disease, assuming that you know nothing about the person’s symptoms or signs?

Your first thought might be:

Well, the test gives a false-positive 5% of the time, so it must be right 95% of the time...so you have a 95% chance of having the disease, right?

Would you be surprised to learn the correct answer is actually only 2%?

Would you be even more surprised (or perhaps alarmed), that when this question was asked 60 house officers, students, and physicians at the Harvard Medical School, that most of them answered this question incorrectly? Almost half of them stated that the answer was 95%. [1](#3znysh7)

This test has been repeated many times amongst medical professionals [2](#2et92p0), law students [3](#tyjcwt), and the general public, all with similar results.

The flaw in reasoning that leads to this mistake is known as the "Base Rate Fallacy", and sometimes (as in your text) as the "False Positive Paradox".

Base Rate Fallacy

A fallacy committed when people attempt to calculate the likelihood of an event in a specific situation, without taking into account the prevalence, or overall likelihood, of that event.

Understanding Bayes' Theorem will help you to avoid this fallacy.

## Reading

You may have done a bit of this reading in your work on Data Exploration 04, but please complete the following:

* Read Section 6.1 - 6.2.2 (pages 247 - 261) of *Machine Learning and Predictive Data Analytics*
* Watch this video about the False Positive Paradox:

1. [Casscells W, Schoenberger A, Graboys TB. Interpretation by physicians of clinical laboratory results. N Engl J Med. 1978;299(18):999-1001. doi:10.1056/NEJM197811022991808](https://www.nejm.org/doi/pdf/10.1056/NEJM197811022991808) [↩](#3dy6vkm)
2. [Whiting PF, Davenport C, Jameson C, et al. How well do health professionals interpret diagnostic information? A systematic review. BMJ Open](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4521525/) [↩](#1t3h5sf)
3. [Two tests of the base rate neglect among law students](https://www.uio.no/studier/emner/jus/jus/JUS4121/v12/undervisningsmateriale/Evidence%20RLE2%20kopi%204%20avd.pdf) [↩](#4d34og8)