

Probabilistic Modeling and Reasoning

The Role of Uncertainty in Machine Learning

Vasilis Gkolemis

ATHENA RC — HUA

June 2025

© Vasilis Gkolemis, 2025. Licensed under CC BY 4.0.

Program

- 1 Course Overview
- 2 The Role of Uncertainty in Machine Learning
- 3 Probabilistic Modeling and Reasoning
- 4 Key Rules of Probability
- 5 Recap

- **Session 1:** Probabilistic Modeling and Reasoning
 - ▶ The Role of Uncertainty in Machine Learning
 - ▶ Probabilistic Modeling and Reasoning
 - ▶ Key Rules of Probability
- **Session 2:** Probabilities and Random Variables
 - ▶ Random Variables and Probability Distributions
 - ▶ Key properties: Expectation, Variance, Covariance
- **Session 3:** Bayesian Modeling: A Unified Framework
 - ▶ Probabilistic vs. Statistical vs. Bayesian Modeling
 - ▶ Prior, Posterior, Likelihood, Predictive Posterior

Course Overview (cont.)

- **Session 4:** Bayesian Linear Regression
 - ▶ Linear Regression as a Probabilistic Model
 - ▶ Exact Inference with Conjugate Priors
- **Session 5:** Bayesian Logistic Regression
 - ▶ Logistic Regression as a Probabilistic Model
 - ▶ Approximate Inference: Laplace Approximation, Importance Sampling, MCMC
- **Session 6:** TODO

Material used in this course

This course is highly inspired by the following material:

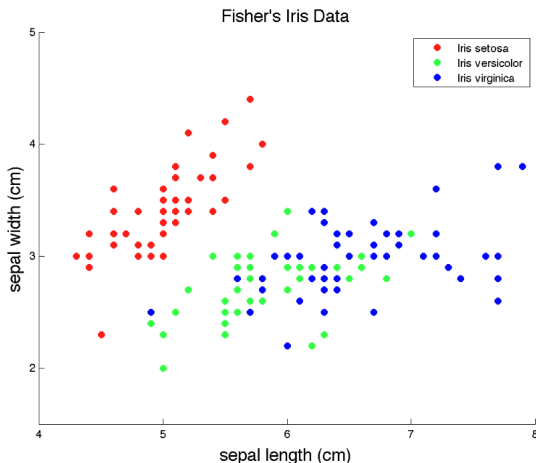
- **Probabilistic Modeling and Reasoning**: University of Edinburgh, School of Informatics,
<https://opencourse.inf.ed.ac.uk/pmr/course-material>
- **Mathematics of Machine Learning**
<https://mml-book.github.io/>
- **BayesRules book**: <https://bayesrulesbook.com/>

Program

- 1 Course Overview
- 2 The Role of Uncertainty in Machine Learning
- 3 Probabilistic Modeling and Reasoning
- 4 Key Rules of Probability
- 5 Recap

Variability

- Variability is part of nature.
- Data for 3 species of iris, from Ronald Fisher (1936).



Variability

- Our handwriting is unique.
- Variability leads to uncertainty: e.g., distinguishing between 1 vs 7 or 4 vs 9.



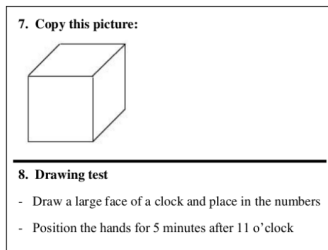
Variability

- Variability leads to uncertainty.
- Reading handwritten text in a foreign language.



Example: Screening and Diagnostic Tests

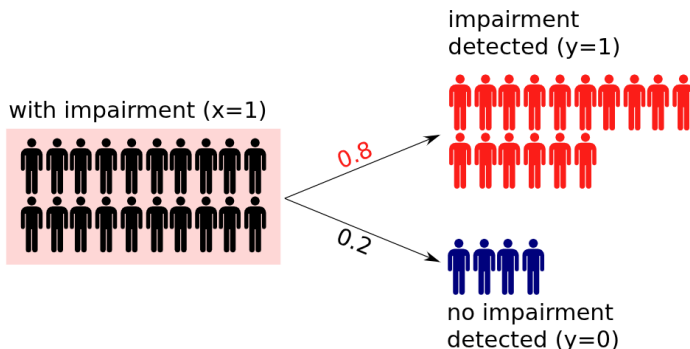
- Early warning test for Alzheimer's disease (Scharre, 2010, 2014).
- Detects “mild cognitive impairment”.
- Takes 10–15 minutes.
- Freely available.
- Assume a 70-year-old man tests positive.
- Should he be concerned?



(Example from sagetest.osu.edu)

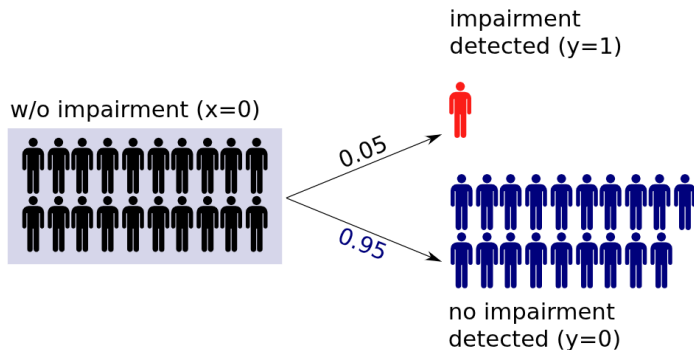
Accuracy of the Test

- Sensitivity of 0.8 and specificity of 0.95 (Scharre, 2010).
- 80% correct for people with impairment.



Accuracy of the Test

- Sensitivity of 0.8 and specificity of 0.95 (Scharre, 2010).
- 95% correct for people without impairment.



Variability Implies Uncertainty

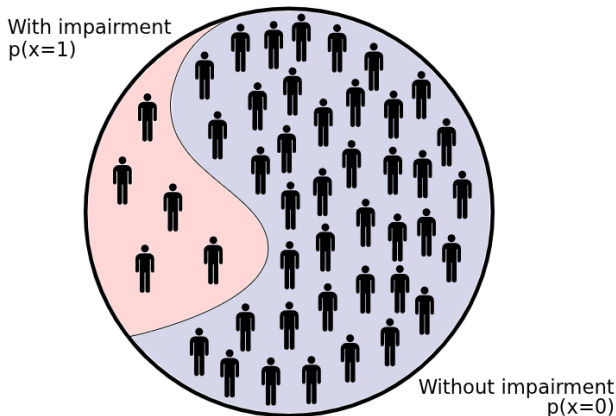
- People of the same group do not have the same test results
 - ▶ Test outcome is subject to variability.
 - ▶ The data are noisy.
- Variability leads to uncertainty.
 - ▶ Positive test \equiv true positive?
 - ▶ Positive test \equiv false positive?
- What can we safely conclude from a positive test result?
- How should we analyze such ambiguous data?

Probabilistic Approach

- $P(y | x)$: model of the test specified in terms of (conditional) probabilities.
- $x \in \{0, 1\}$: quantity of interest (cognitive impairment or not).
- $y \in \{0, 1\}$: test outcome (negative or positive).
- The test outcomes y can be described with probabilities:
 - ▶ Sensitivity = 0.8
 - ★ $\Rightarrow P(y = 1 | x = 1) = 0.8$
 - ★ $\Rightarrow P(y = 0 | x = 1) = 0.2$
 - ▶ Specificity = 0.95
 - ★ $\Rightarrow P(y = 0 | x = 0) = 0.95$
 - ★ $\Rightarrow P(y = 1 | x = 0) = 0.05$

Prior Information

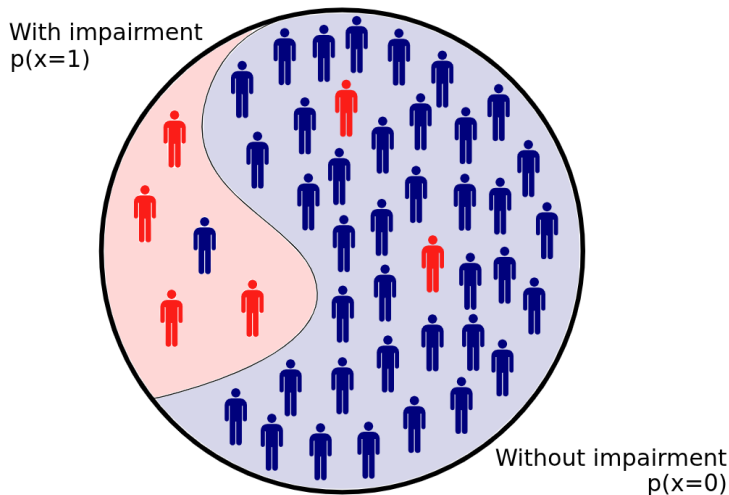
- Among people like the patient, $P(x = 1) = \frac{5}{45} \approx 11\%$ have a cognitive impairment (plausible range: 3% – 22%, Geda, 2014).



Probabilistic Model

- Reality:
 - ▶ Properties/characteristics of the group of people like the patient
 - ▶ Properties/characteristics of the test
- Probabilistic model:
 - ▶ $P(x = 1)$: probability of cognitive impairment
 - ▶ $P(y = 1|x = 1)$ or $P(y = 0|x = 1)$: probability of positive or negative test given cognitive impairment
 - ▶ $P(y = 1|x = 0)$ or $P(y = 0|x = 0)$: probability of positive or negative test given no cognitive impairment
 - ▶ Fully specified by three numbers.
- A probabilistic model is an abstraction of reality that uses probability theory to quantify the chance of uncertain events.

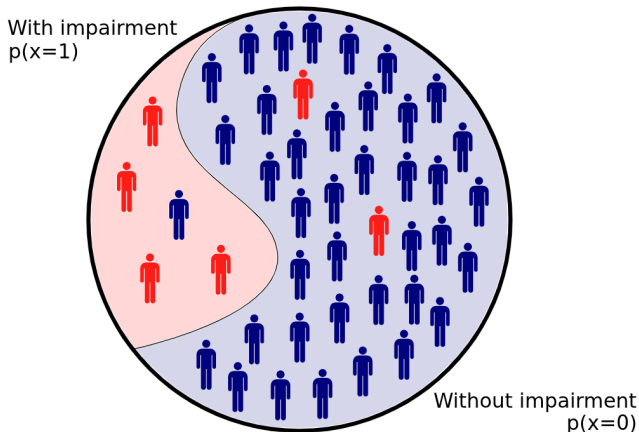
If We Tested the Whole Population



If We Tested the Whole Population

- Fraction of people who are impaired and test positive:

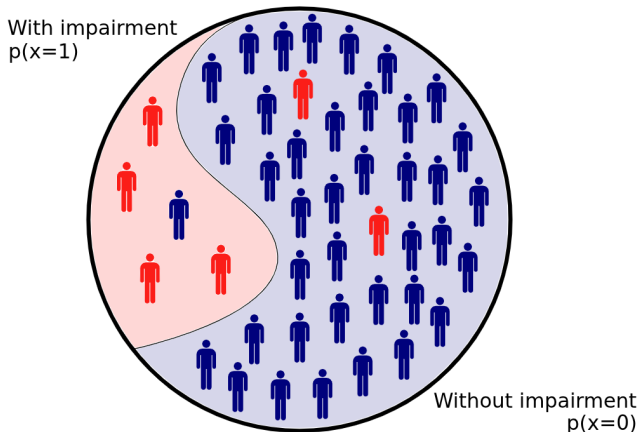
$$P(x = 1, y = 1) = P(y = 1|x = 1)P(x = 1) = 0.8 \cdot \frac{5}{45} = \frac{4}{45} \approx 9\%$$



If We Tested the Whole Population

- Fraction of people who are not impaired and test positive:

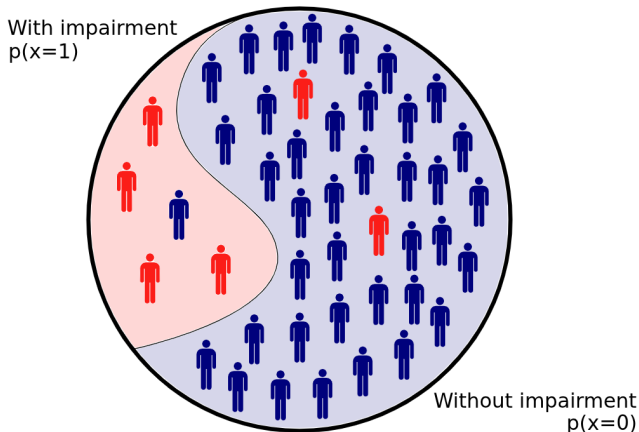
$$P(x = 0, y = 1) = P(y = 1|x = 0)P(x = 0) = 0.05 \cdot \frac{40}{45} = \frac{2}{45} \approx 4\%$$



If We Tested the Whole Population

- Fraction of people where the test is positive:

$$P(y = 1) = P(x = 1, y = 1) + P(x = 0, y = 1) = \frac{4}{45} + \frac{2}{45} = \frac{6}{45} \approx 13\%$$



Putting Everything Together

- Among those with a positive test, fraction with impairment:

$$P(x = 1|y = 1) = \frac{P(y = 1|x = 1)P(x = 1)}{P(y = 1)} = \frac{4}{6} = \frac{2}{3}$$

- Fraction without impairment:

$$P(x = 0|y = 1) = \frac{P(y = 1|x = 0)P(x = 0)}{P(y = 1)} = \frac{2}{6} = \frac{1}{3}$$

- Equations are examples of Bayes' rule.
- Positive test increased probability of cognitive impairment from 11% (prior belief) to 67%, or from 6% to 51%.
- 51% (coin flip)

Program

- 1 Course Overview
- 2 The Role of Uncertainty in Machine Learning
- 3 Probabilistic Modeling and Reasoning**
- 4 Key Rules of Probability
- 5 Recap

- Bayesian analysis is probabilistic modeling and reasoning using Bayes' rule.
- **Probabilistic modeling:**
 - ▶ Model the world (events) using probabilities and random variables.
 - ▶ Example random variables:
 - ★ y : test outcome
 - ★ x : cognitive impairment
- **Probabilistic reasoning (inference):**
 - ▶ Compute probabilities of events given other events.
 - ▶ Infer probabilities of unobserved events from observed data.
 - ▶ Use **Bayes' rule** to update beliefs based on evidence.

- In our example:
 - ▶ Unobserved/uncertain event: cognitive impairment $x = 1$
 - ▶ Observed event (evidence): test result $y = 1$
 - ▶ **Prior**: probability before seeing evidence, e.g., $P(x = 1)$
 - ▶ **Posterior**: updated probability after evidence, e.g., $P(x = 1|y = 1)$
- **Key idea**: The posterior quantifies what we believe about x after seeing the test result y .

Program

- 1 Course Overview
- 2 The Role of Uncertainty in Machine Learning
- 3 Probabilistic Modeling and Reasoning
- 4 Key Rules of Probability**
- 5 Recap

Key Rules of Probability

- **Product rule:**

- ▶ $P(x = 1, y = 1) = P(y = 1|x = 1)P(x = 1)$
- ▶ $P(x = 1, y = 1) = P(x = 1|y = 1)P(y = 1)$

- **Sum rule:**

- ▶ $P(y = 1) = P(x = 1, y = 1) + P(x = 0, y = 1)$

- **Bayes' rule (conditioning) as consequence of product rule:**

- ▶ $P(x = 1|y = 1) = \frac{P(x=1,y=1)}{P(y=1)} = \frac{P(y=1|x=1)P(x=1)}{P(y=1)}$

- **Denominator from sum rule and product rule:**

- ▶ $P(y = 1) = P(y = 1|x = 1)P(x = 1) + P(y = 1|x = 0)P(x = 0)$

Key Rules of Probability

- The rules generalize to multivariate random variables
 - ▶ $\mathbf{x} = (x_1, x_2, \dots)$: vector of random variables
 - ▶ $\mathbf{y} = (y_1, y_2, \dots)$: vector of random variables
- The rules generalize to continuous random variables
- **Product rule:**
 - ▶ $P(\mathbf{x}, \mathbf{y}) = P(\mathbf{y}|\mathbf{x})P(\mathbf{x})$
 - ▶ $P(\mathbf{x}, \mathbf{y}) = P(\mathbf{x}|\mathbf{y})P(\mathbf{y})$
- **Sum rule:**
 - ▶ $P(\mathbf{y}) = \sum_{\mathbf{x}} P(\mathbf{x}, \mathbf{y})$ (discrete case)
 - ▶ $P(\mathbf{y}) = \int P(\mathbf{x}, \mathbf{y}) d\mathbf{x}$ (continuous case)

Program

- 1 Course Overview
- 2 The Role of Uncertainty in Machine Learning
- 3 Probabilistic Modeling and Reasoning
- 4 Key Rules of Probability
- 5 Recap

- **Probabilistic Modeling:**

- ▶ The art of quantifying real-world phenomena with probabilities.
- ▶ Example: Diagnosing Alzheimer's with a medical test.

- **Probabilistic Reasoning (Inference):**

- ▶ Infer probabilities of unobserved events from observed data.

- **Bayesian Analysis:**

- ▶ Bayes' rule updates our beliefs with new evidence.
- ▶ Applied in the Alzheimer's test case.

- **Core Probability Rules:**

- ▶ Product rule, Sum rule, Bayes' rule.
- ▶ These simple rules help us make informed conclusions.

Recap & What's Next

Session 2: Probabilities and Random Variables

- Sample Space and Events
- Random Variables and Probability Distributions
- Key properties: Expectation, Variance, Covariance
- Approximate key properties with Monte Carlo methods