Gaussian mixture models unit:

- Everything you need to know about probability (**today**)
- Lab activity: From histograms to Gaussians
- Mixture models
- Expectation-maximization

Today:

- Overview of unit
- Definition of discrete probability
- Discrete random variables
- Continuous random variables and distributions

- Definition of discrete probability
  - Outcomes, events, probability functions
  - Conditioanl probability
  - The meaning of probability
- Discrete random variables
  - Probability mass functions
  - Expectation and variance
- Continuous random variables and distributions
  - Probability density functions
  - Distributions
  - The Gaussian distribution

Let Ω be a sample space and $\mathcal{A} = \mathscr{P}(\Omega)$ be an event space; A *probability function* $P : \mathcal{A} \to [0, 1]$ fulfills the axioms of probability:

1. For all $A \in \mathcal{A}$, $P(A) \geq 0$.
2. $P(\Omega) = 1$
3. For disjoint sets $A, B \in \mathcal{A}$, $P(A \cup B) = P(A) + P(B)$.

Consider the events

- $P(\{\heartsuit J, \heartsuit Q, \heartsuit K, \diamondsuit J, \ldots \clubsuit K\}) = \frac{12}{52} \approx .23$
- A, the card is red. $P(A) = .5$
- B, the card is a diamond. $P(B) = .25$
- C, the card is a 4. $P(C) = \frac{4}{52} \approx .077$
- D, the card is $\diamondsuit 4$. $P(D) = \frac{1}{52} \approx .019$

Theorem (Bayes's theorem)

If X and Y are events, $P(X|Y) = \dfrac{P(Y|X)P(X)}{P(Y)}$.

Proof.

$$
\begin{aligned}
P(Y|X) &= \frac{P(X \cap Y)}{P(X)} \\
P(X \cap Y) &= P(Y|X)P(X) \\
P(X|Y) &= \frac{P(X \cap Y)}{P(Y)} \\
&= \frac{P(Y|X)P(X)}{P(Y)} \qquad \square
\end{aligned}
$$

In machine learning, we often avoid explicitly referring to the probability space, but instead refer to probabilities on quantities of interest, which we denote by $\mathcal{T}$. In this book we refer to $\mathcal{T}$ as the *target space*.

We introduce a function $X : \Omega \to \mathcal{T}$ that takes an outcome and returns a particular quantity of interest x as a value in $\mathcal{T}$. This association/mapping from Ω to $\mathcal{T}$ is called a *random variable*

The name "random variable" is a great source of misunderstanding as it is neither random nor is it a variable. It is a function.

Deisenroth et al, *Mathematics for Machine Learning*, pg 155

Remark. The target space, that is, the [codomain] $\mathcal{T}$ of the random variable X, us used to indicate the kind of probability space, i.e., a $\mathcal{T}$ random variable. When $\mathcal{T}$ is finite or countably infinite, this is called a discrete random variable. For continuous random variables, we consider only $\mathcal{T} = \mathbb{R}$ or $\mathcal{T} = \mathbb{R}^D$.

ibid, pg 157

Coming up:

Read textbook portions about probability and distributions (due end-of-day Mon, Feb 13)

Do logistic regression assignment (due end-of-day Wed, Feb 15)

Take probability quiz (due class time, Wed Feb 15)

Do reading and response assignment (due end-of-day Fri, Feb 17)

Project dataset confirmation *(due end-of-day Fri, Feb 17)*