Midterm and Final Exams

Take home exams. Can use textbook and slides – no other material!

- **Midterm:**
  - Oct 19-26 (Tuesday – Tuesday).
  - No working groups or TA hours in that week.
  - All questions are sent privately to the TA’s and answered publicly on Piazza
  - Read instructions in the exam document!

- **Final:** Dec 20th 2:00 pm – in person proctored exam.
Classification from Continuous Data

X=1 fire, X=0 no fire

\( F_{Y|X}(y \mid x=0) = \) distribution of smoke (carbon monoxide gas) level when there is no fire.

\( F_{Y|X}(y \mid x=1) = \) distribution of smoke (carbon monoxide gas) level when there is fire.

Given that \( Y=y \), is there a fire? How do we adapt Bayes’ Rule to continuous distributions?

\[
p_{X \mid Y}(x \mid y) = \frac{p_{XY}(x, y)}{p_Y(y)} = \frac{p_{XY}(x, y)}{\sum_{x'} p_{XY}(x', y)}
\]
Bayes’ Rule: Classification from Continuous Data
Continuous Random Variables

- For any discrete random variable, the CDF is **discontinuous and piecewise constant**.
- If the CDF is **monotonically increasing and continuous**, have a continuous random variable:
  \[
  0 \leq F_X(x) \leq 1 \\
  F_X(x_2) \geq F_X(x_1) \text{ if } x_2 > x_1. \\
  \lim_{x \to -\infty} F_X(x) = 0 \hspace{1cm} \lim_{x \to +\infty} F_X(x) = 1
  \]
- The probability that continuous random variable \(X\) lies in the interval \((x_1, x_2]\) is then
  \[
  P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)
  \]
If the CDF is differentiable, its first derivative is called the **probability density function (PDF)**:

\[
f_X(x) = \frac{dF_X(x)}{dx} = F'_X(x)
\]

\[
F_X(x) = \int_{-\infty}^{x} f_X(t) \, dt
\]

By the **fundamental theorem of calculus**:

\[
\int_{x_1}^{x_2} f_X(x) \, dx = F_X(x_2) - F_X(x_1) = P(x_1 < X \leq x_2)
\]

For any valid PDF:

\[
f_X(x) \geq 0
\]

\[
\int_{-\infty}^{+\infty} f_X(x) \, dx = 1
\]
Classification Problems

➢ Which of the 10 digits did a person write by hand?
➢ Is an email spam or not spam (ham)?
➢ Is this image taken in an indoor or outdoor environment?
➢ Is a pedestrian visible from a self-driving car’s camera?
➢ What language is a webpage or document written in?
➢ How many stars would a user rate a movie that they’ve never seen?
Variants of Bayes Rule

\textit{Infer discrete }X\textit{ from discrete }Y:\n
\[ p_{X|Y}(x \mid y) = \frac{p_{X,Y}(x, y)}{p_Y(y)} = \frac{p_X(x)p_{Y|X}(y \mid x)}{p_Y(y)} \]

\[ p_Y(y) = \sum_x p_X(x)p_{Y|X}(y \mid x) \]

\textbf{Example:}
- \( X = 1, 0: \) airplane present/not present
- \( Y = 1, 0: \) something did/did not register on radar

\textit{Infer continuous }X\textit{ from continuous }Y:\n
\[ f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{f_X(x)f_{Y|X}(y \mid x)}{f_Y(y)} \]

\[ f_Y(y) = \int_x f_X(x)f_{Y|X}(y \mid x) \, dx \]

\textbf{Example: }\( X: \) some signal; \textit{“prior” }\( f_X(x)\)
- \( Y: \) noisy version of \( X\)
- \( f_{Y|X}(y \mid x): \) model of the noise

\textit{Infer discrete }X\textit{ from continuous }Y:\n
\[ p_{X|Y}(x \mid y) = \frac{p_X(x)f_{Y|X}(y \mid x)}{f_Y(y)} \]

\[ f_Y(y) = \sum_x p_X(x)f_{Y|X}(y \mid x) \]

\textbf{Example:}
- \( X: \) a discrete signal; \textit{“prior” }\( p_X(x)\)
- \( Y: \) noisy version of \( X\)
- \( f_{Y|X}(y \mid x): \) continuous noise model

\textit{Infer continuous }X\textit{ from discrete }Y:\n
\[ f_{X|Y}(x \mid y) = \frac{f_X(x)p_{Y|X}(y \mid x)}{p_Y(y)} \]

\[ p_Y(y) = \int_x f_X(x)p_{Y|X}(y \mid x) \, dx \]

\textbf{Example:}
- \( X: \) a continuous signal; \textit{“prior” }\( f_X(x)\)
  (e.g., intensity of light beam);
- \( Y: \) discrete r.v. affected by \( X\)
  (e.g., photon count)
- \( p_{Y|X}(y \mid x): \) model of the discrete r.v.
Continuous & Discrete Variables

$X \rightarrow$ discrete random variable with PMF $p_X(x)$

\[ p_X(x) \geq 0, \quad \sum_x p_X(x) = 1. \]

**Example:** Probability of each category in a classification problem.

$Y \rightarrow$ continuous random variable with conditional PDF $f_{Y|X}(y \mid x)$

\[ f_{Y|X}(y \mid x) \geq 0, \quad \int_{-\infty}^{+\infty} f_{Y|X}(y \mid x) \, dy = 1 \quad \text{for all } x. \]

**Example:** Output of temperature sensor, motion sensor, camera, etc. For each possible $X=x$, the distribution of this sensor is different.
Marginal Distributions are Mixtures

Let $X \rightarrow$ discrete random variable with PMF $p_X(x)$

$$p_X(x) \geq 0, \quad \sum_x p_X(x) = 1.$$  

Let $Y \rightarrow$ continuous random variable with conditional PDF $f_{Y|X}(y \mid x)$

$$f_{Y|X}(y \mid x) \geq 0, \quad \int_{-\infty}^{+\infty} f_{Y|X}(y \mid x) \, dy = 1 \quad \text{for all } x.$$

**Marginal Distribution:** Summing over all possible values of $X$, the marginal CDF (PDF) is a *mixture* of the conditional CDFs (PDFs):

$$F_Y(y) = P(Y \leq y) = \sum_x P(X = x \cap Y \leq y) = \sum_x p_X(x) P(Y \leq y \mid X = x) = \sum_x p_X(x) F_{Y|X}(y \mid x).$$

➢ Then differentiating each term in the sum:

$$f_Y(y) = \sum_x p_X(x) f_{Y|X}(y \mid x).$$
Example: Hard Drive Lifetimes

➢ Suppose 90% of hard drives in some laptop computer model have exponentially distributed lifetime param $\theta_0$

$$f_{Y \mid X}(y \mid 0) = \theta_0 e^{-\theta_0 y} \quad p_X(0) = 0.9$$

➢ However, 10% of hard drives have a manufacturing defect that gives them a shorter lifetime $\theta_1 > \theta_0$

$$f_{Y \mid X}(y \mid 1) = \theta_1 e^{-\theta_1 y} \quad p_X(1) = 0.1$$

➢ Recall mean of exponential distribution $Z$:

$$E[Y \mid X = 0] = \frac{1}{\theta_0} > \frac{1}{\theta_1} = E[Y \mid X = 1]$$

$$E[Z] = \int_0^\infty x \theta e^{-\theta x} dx = \left[-xe^{-\theta x} - \frac{1}{\theta} e^{-\theta x}\right]_{x=0} = \frac{1}{\theta}$$
Example: Hard Drive Lifetimes

➢ Suppose 90% of hard drives in some laptop computer model have exponentially distributed lifetime param $\theta_0$

$$f_{Y|X}(y \mid 0) = \theta_0 e^{-\theta_0 y} \quad p_X(0) = 0.9$$

➢ However, 10% of hard drives have a manufacturing defect that gives them a shorter lifetime $\theta_1 > \theta_0$

$$f_{Y|X}(y \mid 1) = \theta_1 e^{-\theta_1 y} \quad p_X(1) = 0.1$$

➢ If your hard drive has operated for $t$ seconds and has not yet failed, what is the probability it is defective?

$$P(X = 1 \mid Y > t) = \frac{P(Y > t \mid X = 1)P(X = 1)}{P(Y > t)}$$

$$= \frac{0.1e^{-\theta_1 t}}{0.1e^{-\theta_1 t} + 0.9e^{-\theta_0 t}}$$

$F_Y(y) = 1 - e^{-\lambda y}$

$f_Y(y) = \lambda e^{-\lambda y}$
Example: Hard Drive Lifetimes

➢ Suppose 90% of hard drives in some laptop computer model have exponentially distributed lifetime param \( \theta_0 \)

\[
f_{Y \mid X}(y \mid 0) = \theta_0 e^{-\theta_0 y} \quad p_X(0) = 0.9
\]

➢ However, 10% of hard drives have a manufacturing defect that gives them a shorter lifetime \( \theta_1 > \theta_0 \)

\[
f_{Y \mid X}(y \mid 1) = \theta_1 e^{-\theta_1 y} \quad p_X(1) = 0.1
\]

➢ If your hard drive fails after exactly \( t \) seconds of operation, what is the probability it is defective?

\[
P(X = 1 \mid Y = t) = \frac{P(Y = t \mid X = 1)P(X = 1)}{P(Y = t)}
\]

Problem: For continuous variable \( Y \), \( P(Y = t) = 0 \)
Discrete Inference from Continuous Data

\( X \rightarrow \text{discrete random variable with PMF } p_X(x) \)

\( Y \rightarrow \text{continuous random variable with conditional PDF } f_{Y|X}(y \mid x) \)

- \( Y \) has a different PDF for each possible discrete \( X=x \)
- Given \( Y=y \), we want to find the probability of each possible \( X=x \)
- But how can we condition on an event of probability zero?

\[
P(X = x \mid Y = y) \approx P(X = x \mid y \leq Y \leq y + \delta)
= \frac{p_X(x)P(y \leq Y \leq y + \delta \mid X = x)}{P(y \leq Y \leq y + \delta)}
\approx \frac{p_X(x)f_{Y|X}(y \mid x)\delta}{f_Y(y)\delta}
= \frac{p_X(x)f_{Y|X}(y \mid x)}{f_Y(y)}.
\]

L'Hôpital's Rule

\[
f_Y(y) = \sum_x p_X(x)f_{Y|X}(y \mid x)
\]
Discrete Inference from Continuous Data

\[ Pr(X = x \mid Y = y) = \lim_{dy \to 0} Pr(X = x \mid y \leq Y \leq y + dy) \]

\[ = \lim_{dy \to 0} \frac{Pr(X = x) Pr(y \leq Y \leq y + dy \mid X = x)}{Pr(y \leq Y \leq y + dy)} \]

L'Hôpital's Rule

\[ = \lim_{dy \to 0} \frac{Pr(X = x) \frac{1}{dy} \int_y^{y+dy} f_Y \mid x(y \mid x) dy}{\frac{1}{dy} \int_y^{y+dy} f_Y(y) dy} \]

\[ = \frac{Pr(X = x) f_Y \mid x(y \mid x)}{f_Y(y)} \]
Example: Hard Drive Lifetimes

➢ Suppose 90% of hard drives in some laptop computer model have exponentially distributed lifetime param \( \theta_0 \)

\[
f_Y | X(y \mid 0) = \theta_0 e^{-\theta_0 y} \quad p_X(0) = 0.9
\]

➢ However, 10% of hard drives have a manufacturing defect that gives them a shorter lifetime \( \theta_1 > \theta_0 \)

\[
f_Y | X(y \mid 1) = \theta_1 e^{-\theta_1 y} \quad p_X(1) = 0.1
\]

➢ If your hard drive fails after exactly \( t \) seconds of operation, what is the probability it is defective?

\[
P(X = 1 \mid Y = t) = \frac{f_Y | X(y \mid 1)p_X(1)}{f_Y(y)}
= \frac{0.1 \theta_1 e^{-\theta_1 t}}{0.1 \theta_1 e^{-\theta_1 t} + 0.9 \theta_0 e^{-\theta_0 t}}
\]
Suppose $X$ is discrete and $Y$ is continuous, and we observe $Y=y$.

**Prior probability mass function for $X$:**

$$p_X(x)$$

**Conditional probability density function for $Y$:**

$$f_{Y|X}(y \mid x)$$

**Posterior probability mass function of $X$ given $Y=y$:**

$$p_{X|Y}(x \mid y) = \frac{p_X(x)f_{Y|X}(y \mid x)}{f_Y(y)}$$

$$f_Y(y) = \sum_x p_X(x)f_{Y|X}(y \mid x)$$
Suppose \( X \) is discrete and \( Y \) is continuous, and we observe \( Y = y \).

**Prior probability mass function for** \( X \):

\[
p_X(x)
\]

**Conditional probability density function for** \( Y \):

\[
f_{Y \mid X}(y \mid x)
\]

**Marginal expected value of random variable** \( Y \):

\[
f_Y(y) = \sum_x p_X(x) f_{Y \mid X}(y \mid x)
\]

\[
E[Y] = \int_{-\infty}^{+\infty} y f_Y(y) \, dy = \sum_x p_X(x) \int_{-\infty}^{+\infty} y f_{Y \mid X}(y \mid x) \, dy
\]

\[
= \sum_x p_X(x) E[Y \mid X = x]
\]
Mixed Continuous/Discrete Distributions

Representations of distribution of $X$:
- CDF is (always) well defined
- Formally, PDF does not exist in this case
- Informally, can illustrate via a hybrid of a probability density function (PDF) and a probability mass function (PMF)
- Related concepts in physics & engineering: Dirac delta function, impulse response

Distribution of random variable $X$:
- With probability 0.5, $X$ has a continuous uniform distribution on $[0,1]$
- With probability 0.5, $X=0.5$