

Business Data Analysis

MGM509

Lecture 8

Reasoning Under Uncertainty:
Bayesian Thinking and Evidence Models

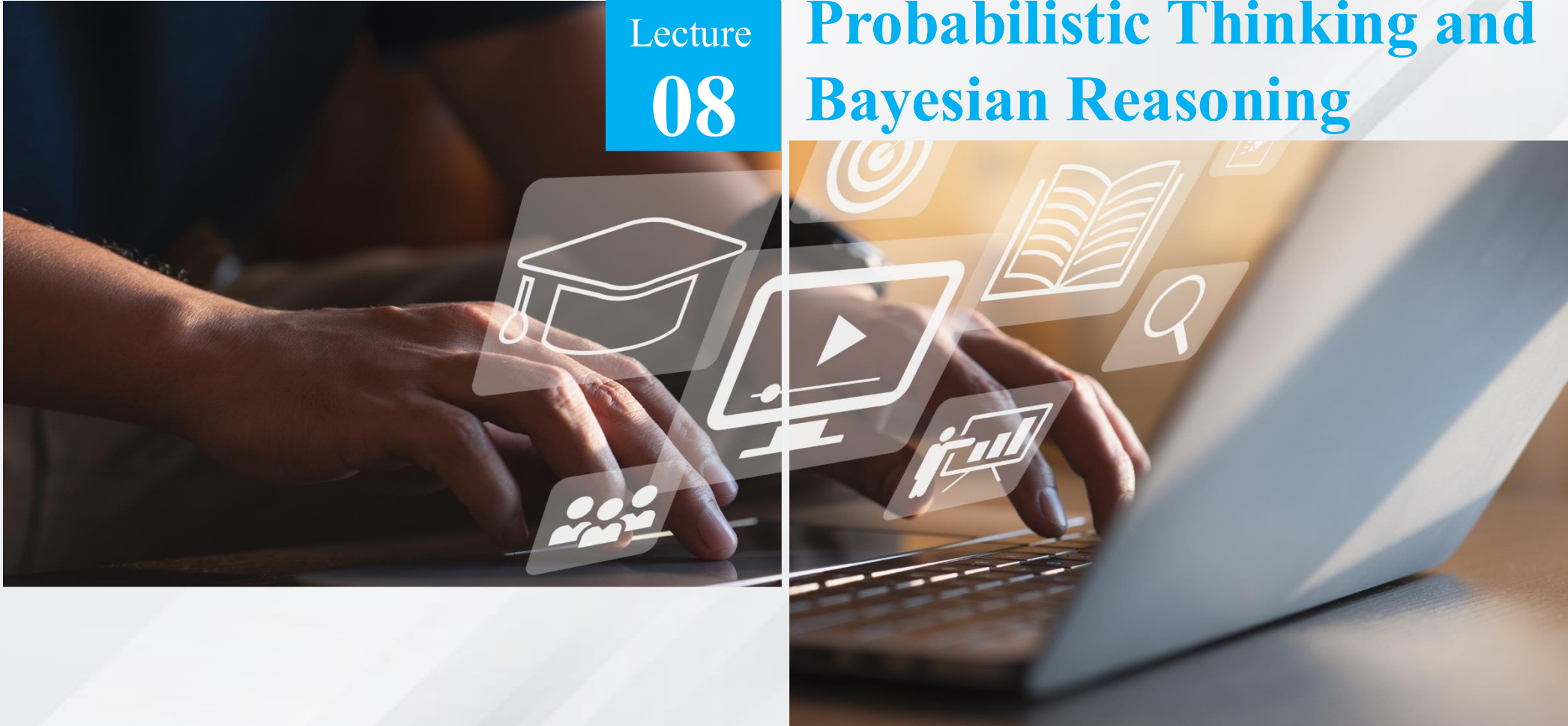
Lecture
8

What we will cover

1. Probabilistic Thinking and Bayesian Reasoning
2. Probabilistic Classification and Decision Thresholds
3. Lift, Ranking, and Model Evaluation Tools
4. Decision Support Using Expected Value and Utility

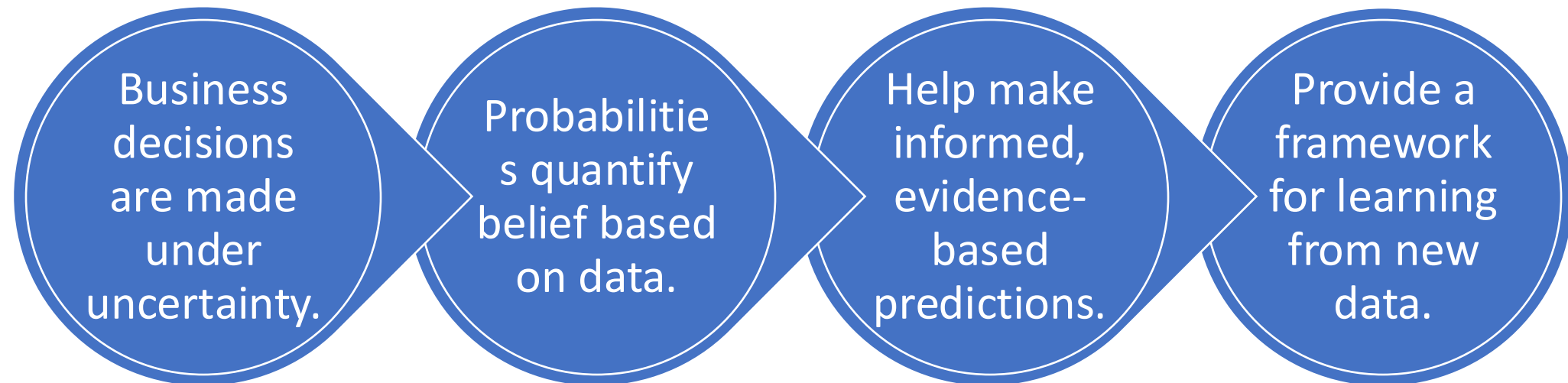
Lecture
08

Probabilistic Thinking and Bayesian Reasoning



**Lecture
8**

Why Probabilities Matter in Data Science?



**Lecture
8****From Data to Decision**

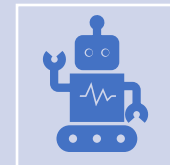
Raw data lacks context until interpreted probabilistically.



Data science models help map evidence to likelihoods.



Supports risk assessment and resource allocation.



Foundation of automated decision-making systems.

**Lecture
8**

The Role of Evidence

Evidence shifts or supports our belief in a hypothesis.

Probabilities enable belief updating over time.

Enables predictive modeling and scenario planning

Examples:

- Predicting churn in telecom and SaaS.
- Fraud detection using transaction patterns.
- Credit risk scoring in financial services.
- Spam filtering based on email features.

**Lecture
8**

What is Bayes Rule?



Developed by Thomas Bayes (1701–1761).



A rule to revise existing beliefs with new evidence.



Core tool in reasoning from incomplete information.



Bridges data and decision-making.

Lecture
8

What is Bayes Rule?

- $$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)}$$

Where:

- $P(H|E)$ = Posterior probability: the probability of hypothesis H given the evidence E
- $P(E|H)$ = Likelihood: the probability of observing evidence E if H is true
- $P(H)$ = Prior probability: the initial belief about hypothesis H before seeing the evidence
- $P(E)$ = Evidence probability: the total probability of observing E under all hypotheses

Bayes' Theorem

$$P(H) = \frac{P(E|H) \cdot P(H)}{P(E)}$$

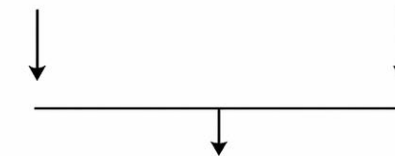
posterior likelihood prior

Hypothesis (H)

Patient has Disease

Evidence (E)

Test is Positive

**Posterior**

Updated probability

**Lecture
8****Understanding Priors and Likelihood****Prior probability**

- Based on expert knowledge or historical data.
- Can be subjective or empirical.
- Strong priors influence conclusions, especially with limited data.
- Priors reduce noise in sparse datasets.

Likelihood

- Measures how expected the evidence is under a hypothesis.
- Crucial for distinguishing between competing explanations.
- Helps assess how data supports a belief.
- Drives the strength of belief revision.

**Lecture
8****Posterior Probability**

Result of applying Bayes Rule.



Guides updated decisions and predictions.



Reflects new evidence combined with existing knowledge.



Used in classifiers, diagnostics, forecasting.

Lecture
8

Bayesian Example: Medical Diagnosis

- Problem: A disease affects 1% of the population. A test for the disease is 99% accurate (true positive rate), but also has a 5% false positive rate.

If someone tests positive, What's the probability they actually have the disease?

Lecture
8

Step-by-step using Bayes' Theorem:

Let:

- H = Patient has the disease
- E = Test result is positive

Known:

- $P(H) = 0.01$ (1% of population has the disease)
- $P(\neg H) = 0.99$ (99% don't)
- $P(E | H) = 0.99$ (true positive rate)
- $P(E | \neg H) = 0.05$ (false positive rate)

Lecture
8

Step-by-step using Bayes' Theorem:

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)}$$

Where:

$$P(E) = P(E|H) \cdot P(H) + P(E|\bar{H}) \cdot P(\bar{H})$$

$$P(E) = (0.99)(0.01) + (0.05)(0.99) = 0.0099 + 0.0495 = 0.0594$$

$$P(H|E) = \frac{0.0099}{0.0594} = \text{Approx. } 0.1667$$

Even after testing positive, the **chance of having the disease is only ~16.7%** due to the rarity of the disease and the false positive rate.

Lecture 8

Bayesian vs Frequentist Thinking

Aspect	Bayesian Thinking	Frequentist Thinking
Definition of Probability	Degree of belief or certainty, subjective	Long-run frequency of events
Prior Knowledge	Explicitly includes prior beliefs (prior probability)	Assumes no prior beliefs; relies solely on sample data
Updates with New Data	Uses Bayes' Theorem to revise beliefs as data accumulates	Data is fixed; probability only comes from long-term repeated trials
Typical Output	Posterior distribution (a probability distribution of outcomes)	Point estimates, confidence intervals, p-values
Interpretation of Parameters	Parameters are random variables (with distributions)	Parameters are fixed constants ; data is random
Use Cases	Small datasets, dynamic models, decision-making under uncertainty	Large datasets, hypothesis testing, classical inference
Example	"There's a 70% chance this customer will churn."	"If we repeated this test 100 times, 95% of intervals would contain the true mean."

Lecture
8

Common Misinterpretations of Probabilities

Base rate fallacy: ignoring prior probabilities.

Confusing $P(A|B)$ with $P(B|A)$.

Misreading conditional probabilities.

Avoid by explicitly modeling priors and evidence.

Lecture
8

Inverse Reasoning

- Inverse reasoning is the process of reasoning from observed evidence back to the most probable cause or explanation.
- It is central to Bayesian inference, where we want to compute:
$$P(\text{Hypothesis} | \text{Evidence})$$
- This is opposite of what most models directly give us (i.e., $P(\text{Evidence} | \text{Hypothesis})$).
- Using Bayes' Rule, we “invert” the known probability to get the one we want.

Lecture
8

Example of Inverse Reasoning

- You observe a customer who hasn't logged in for 3 weeks and downgraded their subscription.
- You want to know: → What is the probability that this customer will churn (leave)?
- This is $P(\text{Churn} \mid \text{Observed Behavior})$ — **inverse reasoning**.
- But your data model gives: $P(\text{Observed Behavior} \mid \text{Churn})$ — how likely a churned customer behaves this way.
- Using **Bayes' Rule**, you compute the **inverted probability**:
$$P(\text{Churn} \mid \text{Behavior}) = P(\text{Behavior} \mid \text{Churn}) \cdot P(\text{Churn}) / P(\text{Behavior})$$

Lecture
08

Probabilistic Classification and Decision Thresholds



Lecture
8

Probabilities in Classification

Classification
predicts
category
labels.

Probabilistic
classifiers
predict
likelihood of
each class.

Allows
flexible
decision
thresholds.

Supports
confidence-
based
predictions.

Lecture
8Email Spam Detection Using
Probabilistic Classification

- The Problem: You receive a new email that contains the word “free”.
- You want to calculate: → What is the probability this email is spam? → That is, $P(\text{Spam} \mid \text{“free”})$

Given From historical data:

- $P(\text{Spam}) = 0.3 \rightarrow 30\%$ of all emails are spam
- $P(\text{“free”} \mid \text{Spam}) = 0.6 \rightarrow 60\%$ of spam emails contain “free”
- $P(\text{“free”} \mid \text{Not Spam}) = 0.1 \rightarrow$ Only 10% of legitimate emails contain “free”
- $P(\text{Not Spam}) = 0.7$

Lecture
8Email Spam Detection Using
Probabilistic Classification

- $P(\text{Spam} \mid \text{free}) = P(\text{free} \mid \text{Spam}) \cdot P(\text{Spam}) / P(\text{free})$. (Bayes' Rule)
- First calculate $P(\text{"free"})$:

$$P(\text{free}) = P(\text{free} \mid \text{Spam}) \cdot P(\text{Spam}) + P(\text{free} \mid \text{Not Spam}) \cdot P(\text{Not Spam})$$

$$P(\text{free}) = (0.6)(0.3) + (0.1)(0.7) = 0.18 + 0.07 = 0.25$$

Now apply Bayes:

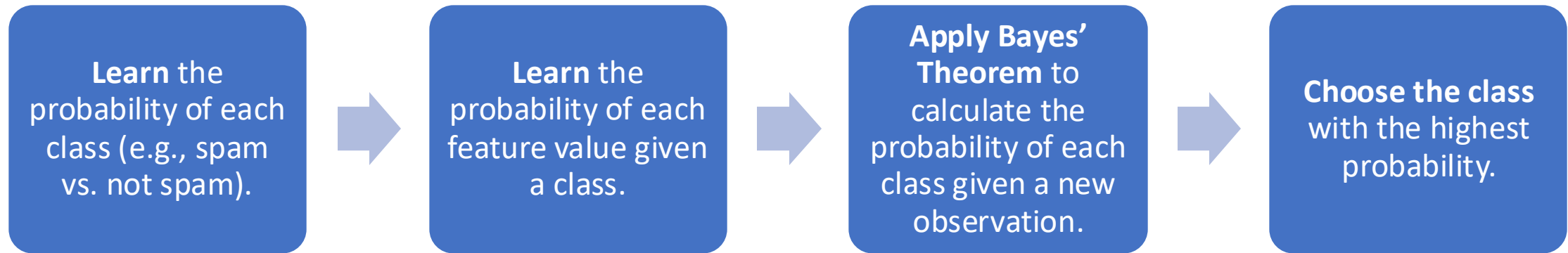
$$P(\text{Spam} \mid \text{free}) = 0.6 \cdot 0.3 / 0.25 = 0.18/0.25 = 0.72$$

There is a 72% probability that the email is spam if it contains the word "free."

Lecture
8

Naive Bayes Classifier

- A Naive Bayes Classifier is a simple and effective probabilistic machine learning algorithm based on Bayes' Theorem.
- It is called “naive” because it assumes that all input features are independent of each other given the class label — an assumption that is rarely true in real-world data, but often works surprisingly well in practice.
- Very efficient for high-dimensional data.
- Popular in text classification.

**Lecture
8****How it works (Simplified)?**

- Example Use Case: In email spam detection, the classifier calculates:
- $P(\text{Spam} \mid \text{Free, "Offer", "Buy"})$ vs. $P(\text{Not Spam} \mid \text{Free, "Offer", "Buy"})$
- It selects the label (spam or not) with the higher probability.

Lecture 8

Why Naive Bayes Works Well?

Feature	Benefit
Fast and scalable	Works well with large datasets, especially text
Simple and interpretable	Easy to understand and implement
Performs well on text data	Excellent for spam detection, sentiment analysis, etc.
Requires little training data	Can work with small datasets effectively
Robust to irrelevant features	Even if independence assumption is violated somewhat

Lecture
8

Applications in Business



Sentiment analysis in social media.



Loan default prediction.



Product recommendation systems.



Customer segmentation.

Lecture 8

Calibration vs. Sharpness

Concept	Measures	Goal
Calibration	How well predicted probabilities match reality	Match predicted confidence to actual outcomes
Sharpness	Concentration of predictions around 0 or 1	Make confident, decisive predictions

Lecture
8

Thresholding & Decision Making

- Thresholding is the process of converting a continuous probability score (output by a classifier) into a binary decision by setting a cutoff value called a threshold.

- Lower threshold = fewer false negatives.
- Higher threshold = fewer false positives.
- Balance depends on cost of each error.

Probability

Threshold = 0.5

Threshold = 0.7

0.82

✓ Positive

✓ Positive

0.65

✓ Positive

✗ Negative

0.45

✗ Negative

✗ Negative

Use Cases:

- Fraud detection: raise threshold to avoid false alarms
- Medical diagnosis: lower threshold to catch all possible positives
- Churn prediction: tune threshold based on customer lifetime value

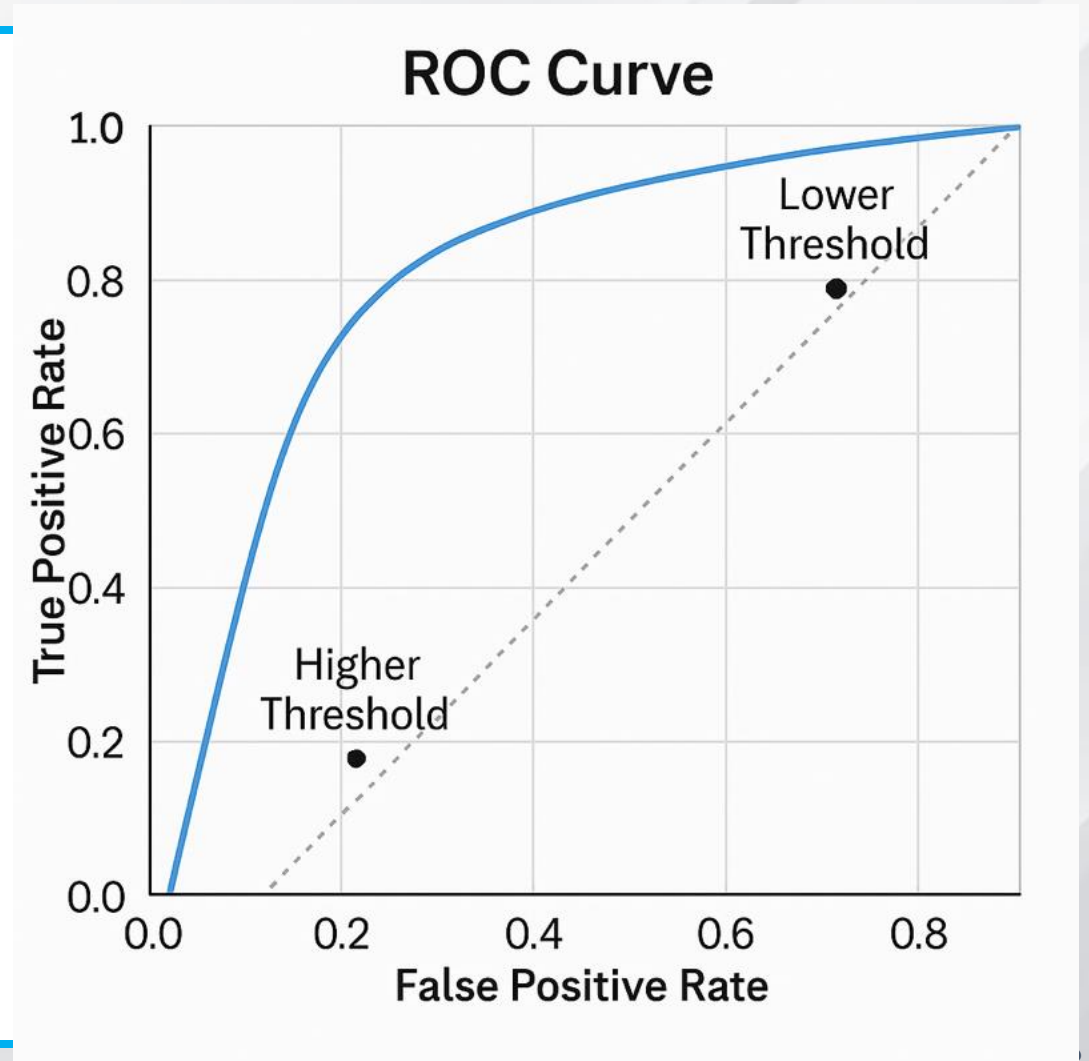
Lecture
8

ROC Curves and Tradeoffs

- **ROC (Receiver Operating Characteristic) Plots:**
 - X-axis:** False Positive Rate (FPR) = $FP / (FP + TN)$
 - Y-axis:** True Positive Rate (TPR) = $TP / (TP + FN)$, also called **Recall**
- TPR vs. FPR at various thresholds.
- Helps visualize model trade-offs.
- Used to compare classifier quality.
- As you adjust the **threshold**, you affect both:
 - True Positives (TP):** Catching real positives
 - False Positives (FP):** Incorrectly flagging negatives as positives

**Lecture
8****Example: ROC Curve**

- Show true positive vs. false positive rate.
- Point closest to top-left = best threshold.



Lecture
8

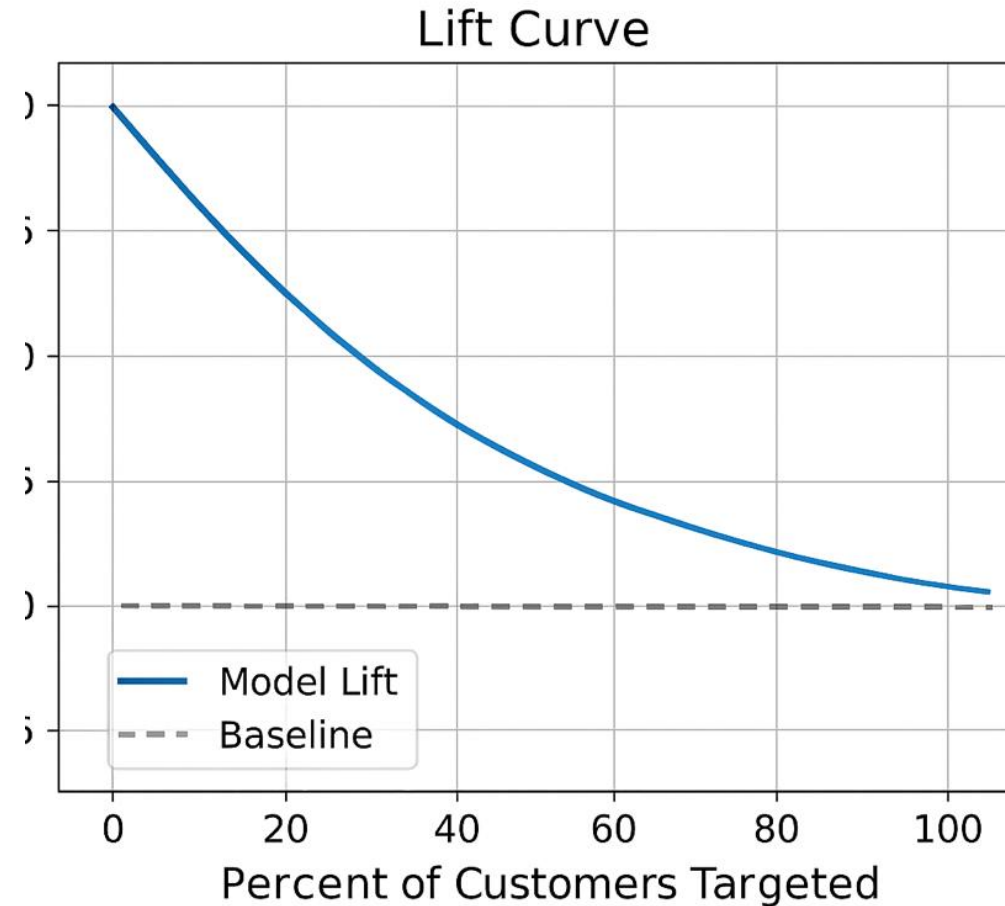
What Is Evidence Lifting?

- **Lift** is a metric that measures how much better a predictive model is at identifying positives compared to random guessing. It's widely used in **classification, marketing, and risk modeling** to evaluate the **effectiveness of a model** or targeting strategy.
- $\text{Lift} = P(\text{target} \mid \text{evidence}) / P(\text{target})$
- Compares conditional vs. unconditional probabilities.
- Used for feature evaluation & ranking.
- Helps rank predictors in modeling.
 - Lift > 1: positive evidence.
 - Lift = 1: neutral.
 - Lift < 1: negative evidence.

**Lecture
8**

Visualizing a Lift Curve

- X-axis: % of population contacted.
- Y-axis: % of total responders captured.
- Steeper curve = better model.
- Include sample chart..



Lecture
8

Common Applications of Lift

Marketing Campaigns & Targeting

- Use Case: Identifying high-value customers likely to respond to a promotion.
- How Lift Helps: Measures how much better the campaign performs by targeting the top decile vs. the general population.
- Example: A lift of 4 in the top 10% means these customers are 4x more likely to respond than

Churn Prediction

- Use Case: Prioritizing customers most likely to leave a service.
- How Lift Helps: Helps allocate retention resources more effectively by targeting customers

Fraud Detection

- Use Case: Identifying transactions likely to be fraudulent.
- How Lift Helps: Lift charts can show how concentrating on top-scoring transactions increases fraud detection efficiency.

Lecture
8

Common Applications of Lift

Credit Scoring and Risk Modeling

- Use Case: Ranking applicants by likelihood of default.
- How Lift Helps: Lift shows how much more likely default is among top-risk groups, aiding in cut-off score decisions.

5. Recommender Systems

- • Use Case: Showing users content/products they're most likely to engage with.
- • How Lift Helps: Measures improvement in click-through or conversion vs. random recommendations.

Lecture
8

Lift vs. ROC Curve

Feature	Lift Curve	ROC Curve
Purpose	Measures model performance relative to random	Evaluates model's ability to discriminate between classes
X-axis	% of population (or customers) sorted by score	False Positive Rate (FPR)
Y-axis	Lift = TPR / Baseline Rate (random targeting)	True Positive Rate (TPR)
Shows	How much better the model is than random guessing in top segments	The trade-off between sensitivity and false positives
Interprets	Model's usefulness for targeting (e.g., top decile)	Model's threshold-independent performance
Best Use Case	When you want to target a top % of ranked list	When you want to evaluate overall classifier quality
Business Focused?	✅ Yes – actionable (target top X%)	❌ Less direct; more technical/diagnostic

Lecture
8

Cautions in Using Lift

Lift can be unstable with rare events.

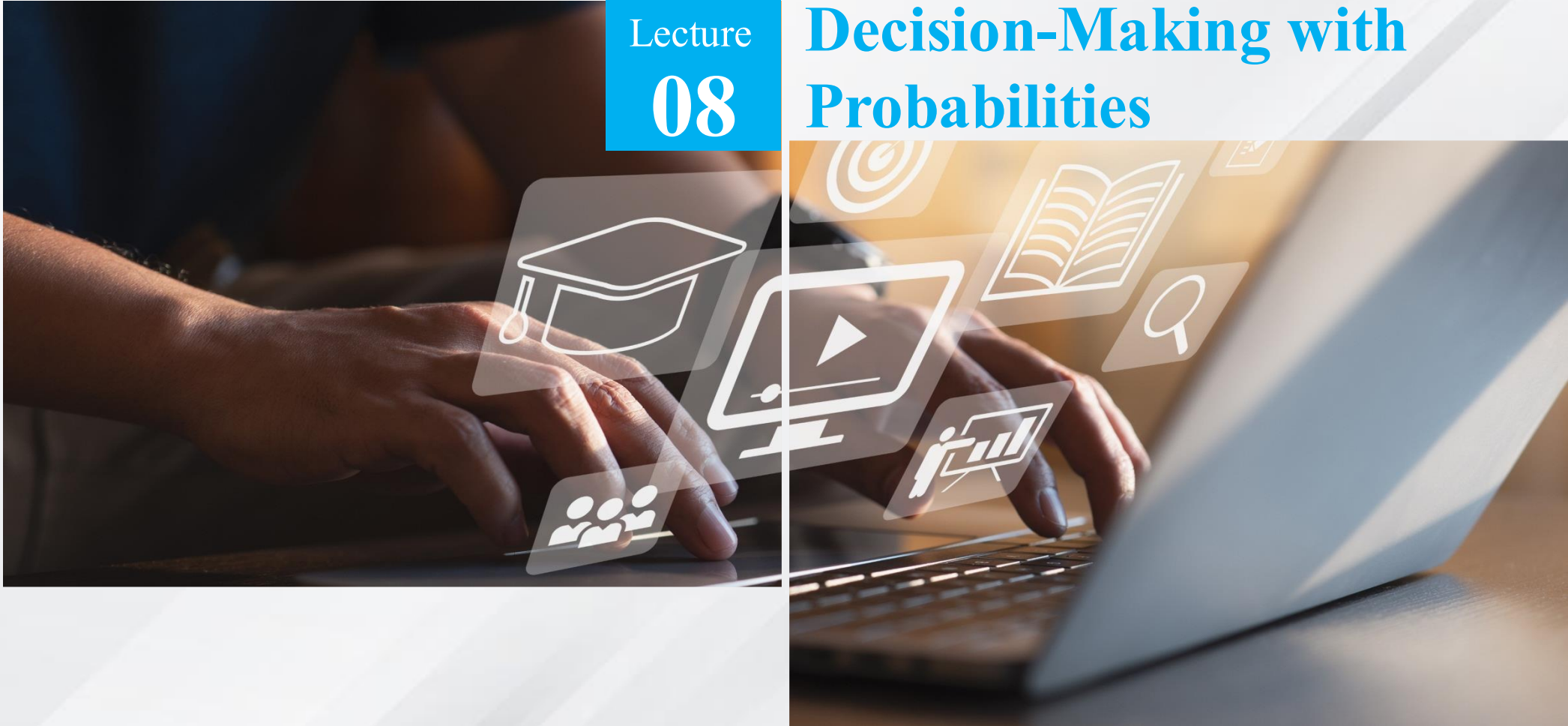
May overfit if not validated.

Validate on holdout or cross-validated sets.

Combine with other metrics for robustness...

Lecture
08

Decision-Making with Probabilities

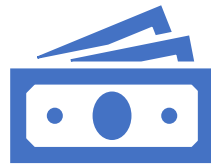


Lecture
8

Decision Making with Probabilities



Use probabilities
as inputs to
decisions.



Estimate expected
utility or value.



Incorporate risk
tolerance and
business goals.



Decision trees
often use this
framework.

Lecture
8

Expected Value and Business Payoff

- The **Expected Value** is the **weighted average** of all possible outcomes, where each outcome is weighted by its probability.
- Expected Value combines model predictions with business impact.
- $EV = \sum \text{Prob}(\text{Event}) \times \text{Value}(\text{Event})$
- Example: $P(\text{sale}) \times \text{Revenue} = \text{Expected Revenue}$
- Choose actions that maximize EV.

Lecture 8

Example: Customer Churn Intervention

Outcome	Probability	Payoff (Value)	Contribution to EV
Correctly saves a customer (TP)	0.4	+\$500	$0.4 \times 500 = +\$200$
Offers discount to loyal customer (FP)	0.2	-\$50	$0.2 \times -50 = -\$10$
Misses actual churning (FN)	0.1	-\$400	$0.1 \times -400 = -\$40$
Correctly ignores non-churner (TN)	0.3	\$0	$0.3 \times 0 = \$0$
Total Expected Value			\$150

This means **each customer decision yields \$150 on average**, justifying the use of the model

Lecture
8

Utility Theory in Decisions

- Utility Theory is a core concept in decision analysis that helps guide rational decision-making under uncertainty. It shifts the focus from just outcomes to what those outcomes mean to the decision-maker — i.e., how much value (or utility) they bring.
- What Is Utility?
 - Utility is a numerical measure of preference or satisfaction that a person (or business) assigns to an outcome.
 - It reflects how desirable or valuable an outcome is, beyond just dollars — incorporating risk tolerance, context, and personal/business goals.
- $EU = \sum P(\text{outcome}) \cdot U(\text{outcome})$

Lecture 8

Utility vs. Monetary Value

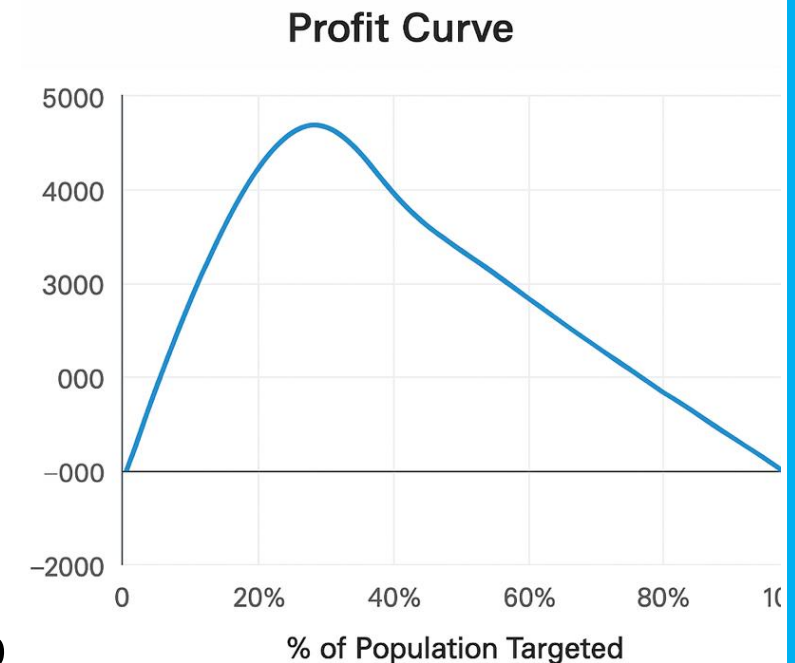
- Two decisions may have the same expected monetary value, but different utilities, especially when risk is involved.
- Example:
 - \$1000 guaranteed = Utility = High
 - 50% chance of \$2000 or \$0 = Same expected value, but lower utility if you're risk-averse

Scenario	How Utility Helps
Marketing	Target customers based on expected utility, not just conversion
Credit Risk	Approve loans that align with risk-adjusted utility, not just profit
Medical AI	Recommend treatments considering patient preferences and risks
Customer Retention	Offer incentives where the utility of keeping a customer outweighs the cost
Fraud Detection	Balance false positives vs. false negatives based on utility of outcomes

Lecture
8

Profit Curves

- A profit curve shows how much profit a classification model generates as a function of how many top-ranked predictions (e.g., customers) you act upon.
- X-axis: % of the population targeted (e.g., top 10%, 20%, ..., 100%)
- Y-axis: Profit or expected value (e.g., revenue gained – cost incurred)
- Unlike ROC or precision-recall curves, profit curves:
 - Integrate domain-specific cost-benefit values
 - Answers: “How much money will this model help us make (or save)?”
 - Help choose the optimal cutoff point



Lecture
8

Example: Churn Campaign Profit Curve

% Targeted	True Positives	False Positives	Profit
10%	80	20	\$3,500
20%	140	60	\$5,000
100%	200	300	-\$2,000

→ Best profit at **20% targeting level**: Model helps **maximize ROI** by focusing only on top scores.

Lecture
8

Key Points

Probabilistic Models Provide Confidence, Not Certainty

Unlike simple classifications, probabilistic models output **likelihoods**, enabling nuanced decisions based on uncertainty and expected outcomes.

Bayes' Theorem Enables Inverse Reasoning

Bayes' Rule updates prior beliefs using observed evidence. It's essential for reasoning from **effect to cause** — a common task in diagnostics, fraud, and risk modeling.

Naive Bayes Classifier Assumes Feature Independence

Despite its “naive” assumption, Naive Bayes performs well in many real-world scenarios (like spam detection and sentiment analysis) due to its simplicity and speed.

Calibration and Expected Value Connect to Business Decisions

A well-calibrated model aligns predicted probabilities with actual outcomes. **Expected Value (EV)** quantifies the payoff of acting on predictions, guiding cost-benefit decision-making.

Bayesian Thinking Enhances Decision Quality Under Uncertainty

By incorporating prior knowledge and evidence, Bayesian reasoning supports better choices — especially in **low-data** or **high-risk** domains like medicine, finance, and cybersecurity.

For Discussion Panel and Next Lecture:

- Chapter 9 “Evidence and Probabilities”

**Lecture
8**



**For Discussion Panel
and Next Lecture**



Thank You