Probability & Estimation

INSTRUCTOR: HONGJIE CHEN

MAY 26TH 2022

A Probabilistic Perspective

- In supervised function approximation
  - instead of learning a function $f : X \rightarrow Y$
  - Learn $P(Y \mid X)$

- Recall **Supervised Learning**: learning from labeled training data

Review Notations in Probability

- Random Variables in capital letter A, or R.V.
- values in lowercase letter a, $P(A = a)$, or $P(a)$ for shorthand

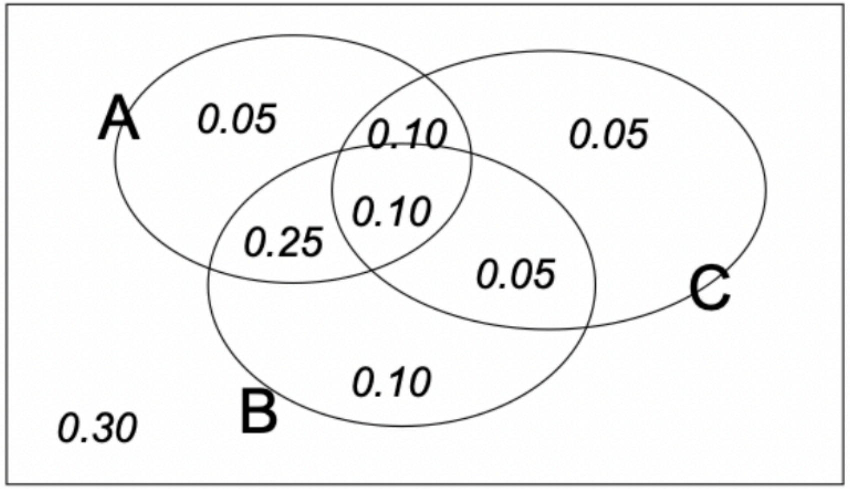
- $P(A \mid B)$, Conditional probability
- $P(A, B)$, Joint probability

- $P(AB) = P(A)P(B)$, independence
- $P(AB \mid C) = P(A \mid C)P(B \mid C)$, conditional independence

Joint Probability Distribution

- Steps for coming up with a joint distribution
  - Make a table listing all combinations of values of R.V.
  - Assign proabability for each combination
  - By axioms of probability, all probability values sum to 1

A	B	C	Prob
0	0	0	0.30
0	0	1	0.05
0	1	0	0.10
0	1	1	0.05
1	0	0	0.05
1	0	1	0.10
1	1	0	0.25
1	1	1	0.10



Using the Joint Probability Distribution

- Now that we have the distribution, we can calculate probability of
  - $P(A)$
  - $P(AB)$
  - $P(A|B)$

A	B	C	Prob
0	0	0	0.30
0	0	1	0.05
0	1	0	0.10
0	1	1	0.05
1	0	0	0.05
1	0	1	0.10
1	1	0	0.25
1	1	1	0.10

Inference with Joint Probability Distribution

- Suppose we want to learn the function $f(G, H) \rightarrow W$ or the probability distribution $P(W \mid G, H)$ of the given data

gender	hours_worked	wealth	
Female	v0:40.5-	poor	0.253122
		rich	0.0245895
	v1:40.5+	poor	0.0421768
		rich	0.0116293
Male	v0:40.5-	poor	0.331313
		rich	0.0971295
	v1:40.5+	poor	0.134106
		rich	0.105933

- Calculate $P(W = \text{rich} \mid G = female, H = \text{40.5-})$
- Can we solve $P(Y \mid X)$ similarly? What do we need?

Figure credit: [link]

Exponential Growth of Table

- Learning $P(Y|X)$ requires all combinations of all values of all random variables.

- Regard a joint probability distrbution with 50 boolean features
  - How many rows?
  - Fraction of rows with zero training samples?

- **Data Sparsity**