CSE 840: Computational Foundations of Artificial Intelligence November 15, 2023

Bayes Theorem to correlation (discrete case)

Instructor: Vishnu Boddeti *Scribe: Avi Lochab, Keerthi Gogineni, Varuntej Kodandapuram*

1 Bayes Theorem

1.1 Law of total probability:

Let $B_1, B_2...., B_k$ be a disjoint partition of Ω with $B_i \in \mathcal{A}$ for all i, and $A \in \mathcal{A}$. Then

$P(A) = \sum_{i=1}^{k} P(A|B_i) \cdot P(B_i) = \sum_{i=1}^{k} P(A \bigcap B_i)$

1.2 Bayes Formula:

$P(B_i|A) = \frac{P(A|B_i) \cdot P(B_i)}{\sum_i P(A|B_i) \cdot P(B_i)} = \frac{P(A \bigcap B_i)}{P(A)}$
Eg: prob(poor reviews|accept), P(accept|poor reviews)

1.3 Example:

COVID Testing

- Assume that 1% of all humans have COVID.
- 90% of people with COVID test positive: "True Positive"
- 8% of people without COVID test positive: "False Positive"
- Given that a person tested positive, What is the likelihood that they have COVID?

$P(COVID|+vetest) = \frac{P(+ve\ test|COVID).P(COVID)}{P(+ve|COVID).P(COVID)+P(-ve|COVID).P(not\ COVID)}$

$= \frac{0.9\times 0.001}{0.9\times 0.01+0.08\times 0.99}$
$\approx 10\%$

2 Independence

2.1 Def:

Consider a probability space $(\Omega, \mathcal{A}, P)$.Two events A,B are called independent if

$$P(A \bigcap B) = P(A) \cdot P(B)$$

Observation: A is independent of B $\leftrightarrow P(A|B) = P(A)$

A family of events $(A_i)_{i\in I}$ is called independent if for all finite subsets J $\subseteq I$ we have

$$P(\bigcap_{i\in J} A_i) = \pi_{i\in J} P(A_i)$$

(Family is called pairwise independent if $\forall i, j \in I$: $P(A_i \bigcap A_j) = P(A_i) \cdot P(A_j)$. This does not imply that the family of events is independent)

2.2 Def:

Two random variables $X : \Omega \to \Omega_1, Y : \Omega \to \Omega_2$ are called independent if their induced σ- algebras $\sigma(X)$, $\sigma(Y)$ are independent: $\forall A \in \sigma(X), B \in \sigma(Y) : P(A \bigcap B) = P(A) \cdot P(B)$.

Notation for independence

- (events) $A \perp\!\!\!\perp B$
- (RV) $X \perp\!\!\!\perp Y$

$A \perp B$ *$X \perp Y$* *Don't do this* *This is orthogonal*

Key concept:

- **Prob:** Central Limit Theorem (CLT), sum of independent RVs.

- **Fairness:** $\hat{Y} \perp\!\!\!\perp S$ enforce while learning.
- **Learning theory:** Assumption that samples are independent w.r.t. each other.

3 Expectation (discrete case)

Consider a discrete random variable $X : \Omega \to \mathbb{R}$ (that is, the image $X(\Omega)$ is at most countable).

3.1 Def:

Let $(\Omega, \mathcal{A}, P)$ be a probability space, $S \subset \mathbb{R}$ at most countable, $X : \Omega \to S$ random variable.

If $\sum_{r \in S} |r| \cdot P(X = r) < \infty$, then

$$E(X) := \sum_{r \in S} r \cdot P(X = r)$$

is called the expectation of X.

(Sometimes written as $E[X]$, $\mathbb{E}X$ or $\mathbb{E}(X)$)

Examples:

- **Toss a coin.** Let $\Omega = \{\text{head}, \text{tail}\}$, $A = P(\Omega)$

$$P(\text{head}) = p, \quad P(\text{tail}) = 1 - p. \quad 0 < p < 1$$
$$X : \Omega \to \{0, 1\}, \quad \text{head} \mapsto 0, \quad \text{tail} \mapsto 1.$$
$$\begin{aligned} E(X) &= 0 \cdot P(X = 0) + 1 \cdot P(X = 1) \\ &= p \cdot 0 + (1 - p) \cdot 1 \\ &= 1 - p \end{aligned}$$

- **Test error of a classifier.**

$$\hat{Y} = f(x) \text{ where } x \text{ is input, } f(\cdot) \text{ classifier}$$

Y is the output.

$$e = (\hat{Y} - Y)^2 = (f(x) - Y)^2$$
$$\min_f E_x(e)$$

3.2 Def:

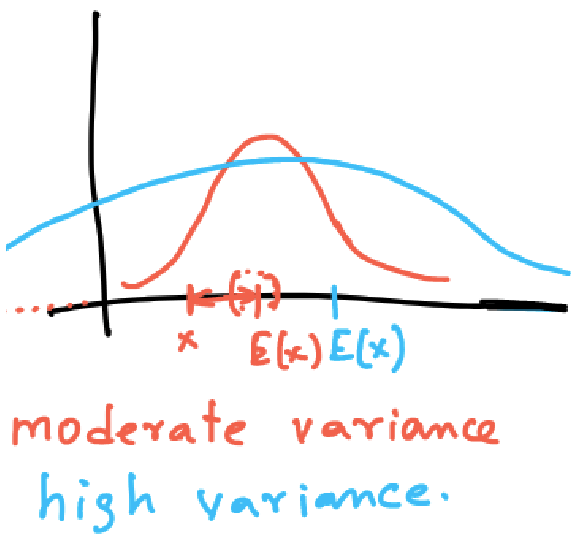
A random variable is called "centered" if $E(X) = 0$.

Important properties:

- Linear: $E(a \cdot X + b \cdot Y) = a \cdot E(X) + b \cdot E(Y)$ where $a, b \in \mathbb{R}$ and X, Y are random variables.
  $a, b \Rightarrow \mathbb{R}$
  $X, Y \Rightarrow RV$
- If X, Y are independent $\Rightarrow E(X \cdot Y) = E(X) \cdot E(Y)$.

4 Variance and Standard Deviation

X, Y: (Ω, A, P) -> $\mathbb{R}$ discrete random variables with $E(x^2) < \infty$, $E(y^2) < \infty$
Then, $Var(x) := E((x - E(x))^2)$



$\sqrt{Var(x)} =: \sigma_x$ is called Standard Deviation

Covariance
Cov(X, Y) := E((X - E(X) * (Y - E(Y))

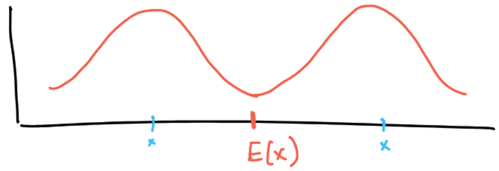
Correlation coefficient =

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y} \in [-1, 1]$$

If Cov(X, Y) = 0, then X and Y are uncorrelated.

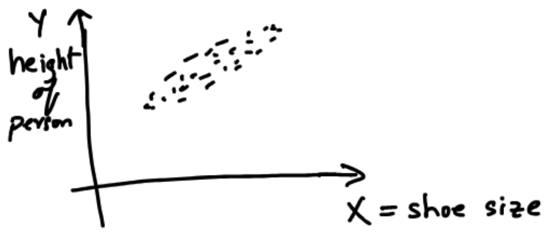
Additionally, for k $\in \mathbb{N}$ we define the terms E(X^k) ("k-th moment")

$E((X - E(X))^k)$ ("k-th centered moment")



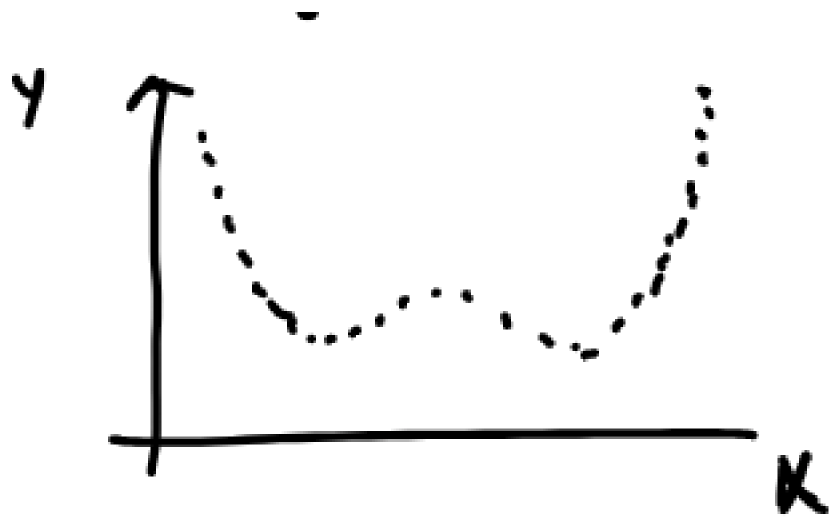
Intuition about covariance

Cov(X,Y) = E(((X - E(X) * (Y - E(Y)))



Positive large covariance

Negative large covariance



uncorrelated

Independence implies uncorrelated BUT uncorrelated does not imply independence

Independence

Properties:
$Var(x) = E(x^2) - (E(x))^2$
$Cov(X, Y) = E(X * Y) - E(X) * E(Y)$ and if $(X \perp Y, E(X) * E(Y) = E(X * Y))$
$E(aX + b) = a * E(X) + b$
$Var(aX + b) = a^2 * Var(X)$
$Cov(X, Y) = Cov(Y, X)$
$Var(X, Y) = Var(X) + Var(Y) + Cov(X, Y)$
If X, Y are independent, Cov(X, Y) = 0 and Var(X, Y) = Var(X) + Var(Y)