CS-XXX: Introduction to Machine Learning

Probability

Arbish Akram

Department of Computer Science
Government College University

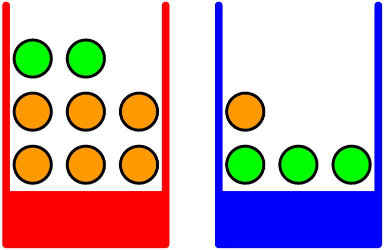
Slides by Dr. Nazar Khan

Probability Theory

- **Uncertainty** is a key concept in pattern recognition.
- Uncertainty arises due to
  - Noise on measurements.
  - Finite size of data sets.
- Uncertainty can be **quantified** via probability theory.

Probability

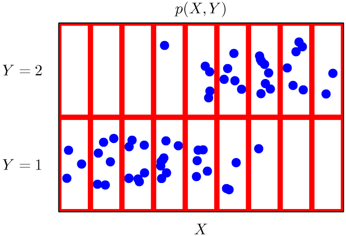
- P(event) is fraction of times event occurs out of total number of trials.
- $P = \lim_{N\to\infty} \frac{\#successes}{N}$.



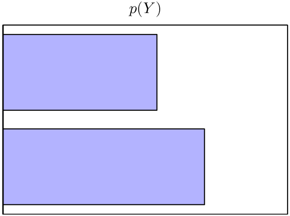
$$P(B = b) = 0.6$$
$$P(B = r) = 0.4$$
$$p(apple) = p(F = a) = ?$$
$$p(\text{blue box given that apple was selected}) = p(B = b|F = a) = ?$$

Terminology

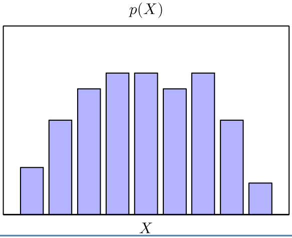
Joint Probability



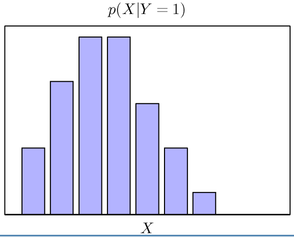
Marginal Probability



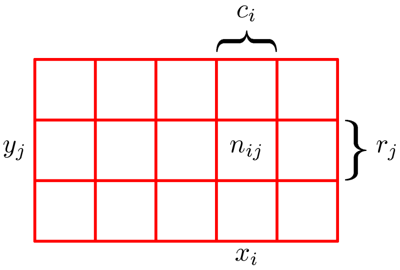
Marginal Probability



Conditional Probability



Elementary rules of probability



Elementary rules of probability

- Sum rule: $p(X) = \sum_Y p(X, Y)$
- Product rule: $p(X, Y) = p(Y|X)p(X)$

These two simple rules form the basis of *all* the probabilistic machinery that will be used in this course.

- The sum and product rules can be combined to write

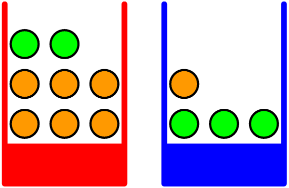
$$p(X) = \sum_Y p(X|Y)p(Y)$$

- A fancy name for this is **Theorem of Total Probability**.
- Since $p(X, Y) = p(Y, X)$, we can use the product rule to write another very simple rule

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$

- Fancy name is **Bayes' Theorem**.
- Plays an *important role* in machine learning.

Terminology

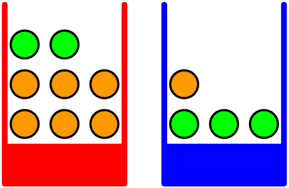


$$P(B = r) = 0.4$$

$$P(B = b) = 0.6$$

- If you don't know which fruit was selected, and I ask you which box was selected, what will your answer be?
  - The box with greater probability of being selected.
  - Blue box because $P(B = b) = 0.6$.
  - This probability is called the **prior probability**.
  - Prior because the data has not been observed yet.

Terminology



$$P(B = r) = 0.4$$
$$P(B = b) = 0.6$$

- Which box was chosen given that the selected fruit was orange?
  - The box with greater $p(B|F = o)$ (via Bayes' theorem).
  - Red box
  - This is called the **posterior probability**.
  - Posterior because the data has been observed.

Independence

- If random variable X is **independent** of random variable Y, then

$$P(X = x|Y = y) = P(X = x)$$

  for all values x and y.

- Then, by the product rule

$$P(X, Y) = P(X|Y)P(Y) = P(X)P(Y)$$

- If joint $p(X = x, Y = y)$ equals the product of marginals $p(X = x)p(Y = y)$ *for all values x and y*, then random variables X and Y are **independent**.
- Intuitively, if Y is independent of X, then knowing X does not change the chances of Y and vice versa.
- Example: if fraction of apples and oranges is same in both boxes, then knowing which box was selected does not change the chance of selecting an apple.

Probability density
Continuous

- So far, our set of events was discrete.
- Probability can also be defined for continuous variables via

$$\text{Prob}(x \in (a, b)) = \int_a^b p(x)dx$$

- *Probability density function* $p(x)$
  - is always non-negative, and
  - integrates to 1.

> Caution: Probability density is not the same as probability. Density can be greater than 1.

Probability density

Continuous

- Sum rule: $p(x) = \int p(x,y)dy$.
- Product rule: $p(x,y) = p(y|x)p(x)$
- Probability density can also be defined for a multivariate random variable $\mathbf{x} = (x_1, \dots, x_D)$.

$$p(\mathbf{x}) \geq 0$$

$$\int_{\mathbf{x}} p(\mathbf{x})d\mathbf{x} = \int_{x_D} \cdots \int_{x_1} p(x_1, \dots, x_D)dx_1 \dots dx_D = 1$$