CS722/822: Machine Learning

Instructor: Jiangwen Sun
Computer Science Department

Probabilistic Theory

Basic Concepts

- An **experiment** (**random variable**) is a well-defined process with observable outcomes.

- The set or collection of all outcomes of an experiment is called the **sample space**, S.

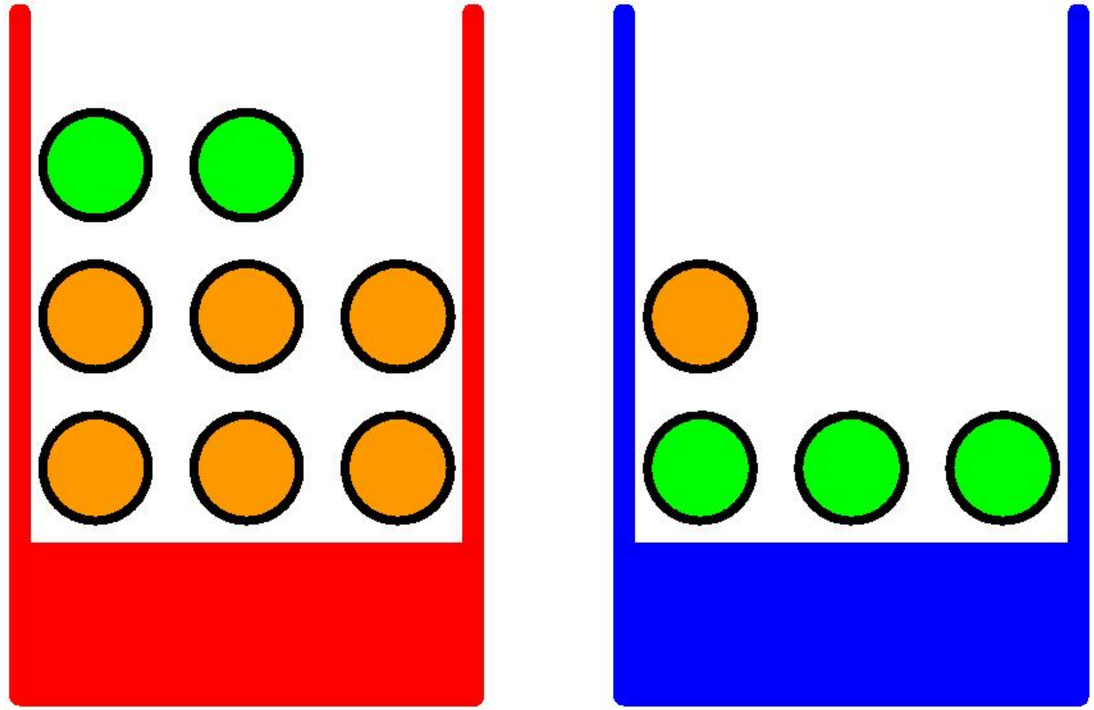
- An **event** E is any subset of outcomes from S.

- **Probability of an event**, P(E) is P(E) = number of outcomes in E / number of outcomes in S.

An Example

Apples and Oranges

X: identity of the fruit
Y: identity of the box



Assume:
P(Y=r) = 40%, P(Y=b) = 60% (prior)

P(X=a|Y=r) = 2/8 = 25%
P(X=o|Y=r) = 6/8 = 75%

(Conditional)

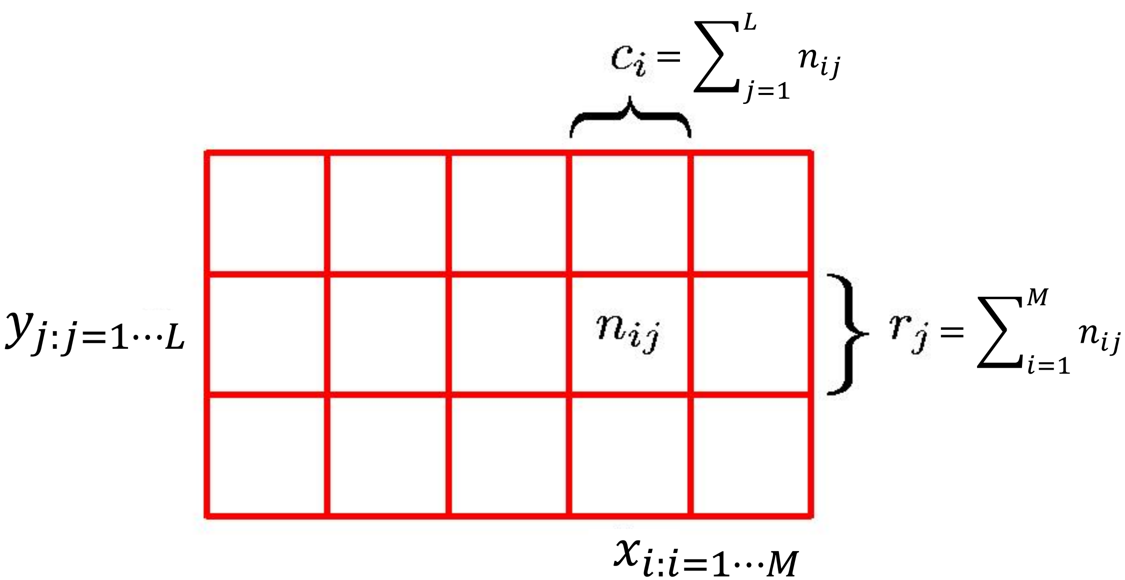
P(X=a|Y=b) = 3/4 = 75%
P(X=o|Y=b) = 1/4 = 25%

Marginal
P(X=a) = ?
P(X=o) = ?

Posterior
P(Y=r|X=o) = ?
P(Y=b|X=o) = ?

A More General Case

Two random variables, X and Y



Contingency Table

- Joint Probability

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

- Marginal Probability

$$p(X = x_i) = \frac{c_i}{N}$$

- Conditional Probability

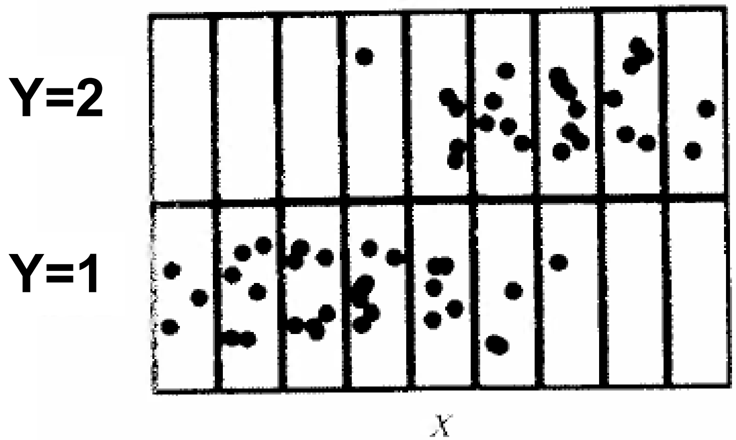
$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$

The total number of incidences is assumed to be N

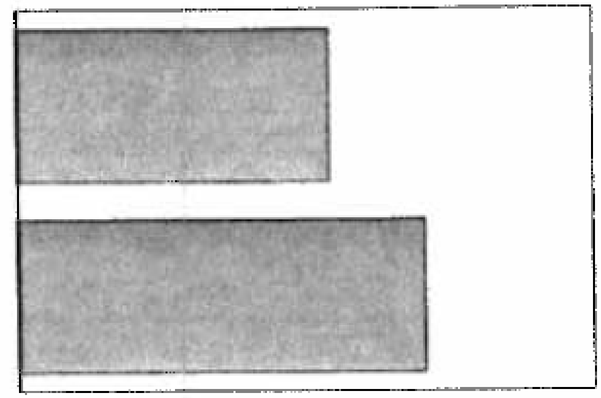
Illustration

$N = 60$

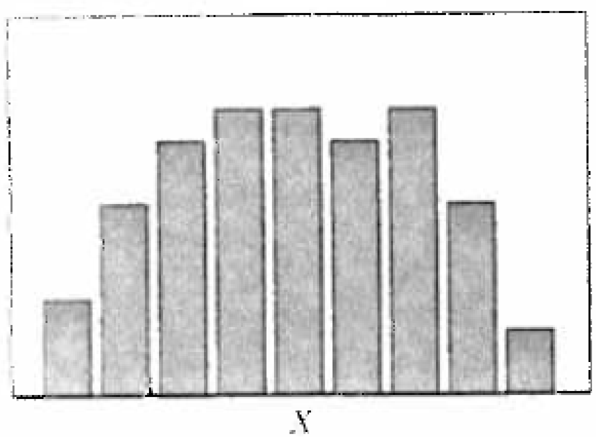
***p*(X,Y)**



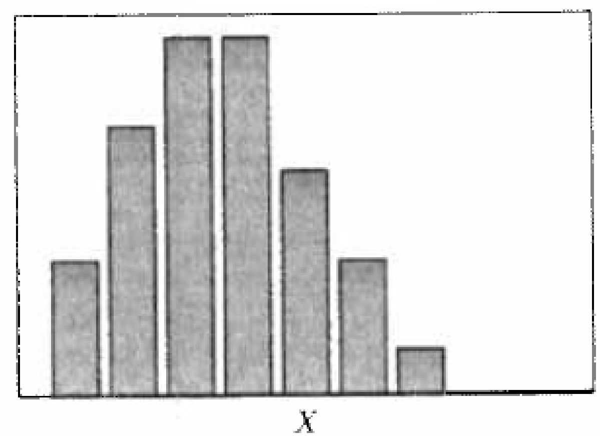
***p*(Y)**



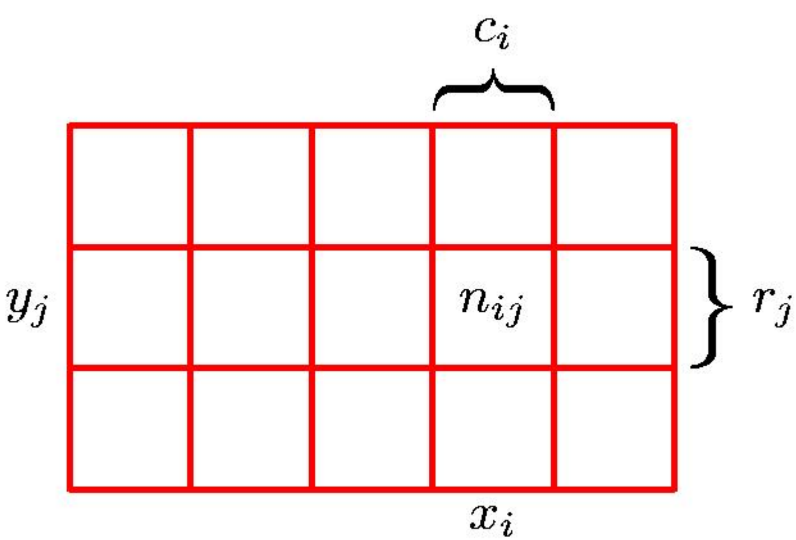
***p*(X)**



***p*(X|Y=1)**



Two Basic Rules of Probability



- **Sum Rule**

$$p(X = x_i) = \frac{c_i}{N} = \frac{1}{N}\sum_{j=1}^{L} n_{ij}$$

$$= \sum_{j=1}^{L} p\left(X = x_i, Y = y_j\right)$$

The marginal prob of X equals the sum of the joint prob of X and Y with respect to Y

- **Product Rule**

$$p\left(X = x_i, Y = y_j\right) = \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \times \frac{c_i}{N} = p\left(Y = y_j | X = x_i\right)p(X = x_i)$$

The joint prob of X and Y equals the product of the conditional prob of Y given X and the marginal prob of X

Two Basic Rules of Probability
-- Compact Notation

- The probability of X:

$$p(X)$$

- The probability of X and Y:

$$p(X,Y)$$

- The probability of Y given X:

$$p(X|Y)$$

- **Sum Rule:**

$$p(X) = \sum_{Y} p(X,Y)$$

- **Product Rule:**

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

Bayes' Rule

- **Sum Rule:**

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

- **Product Rule:**

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

- **Bayes' Rule**

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

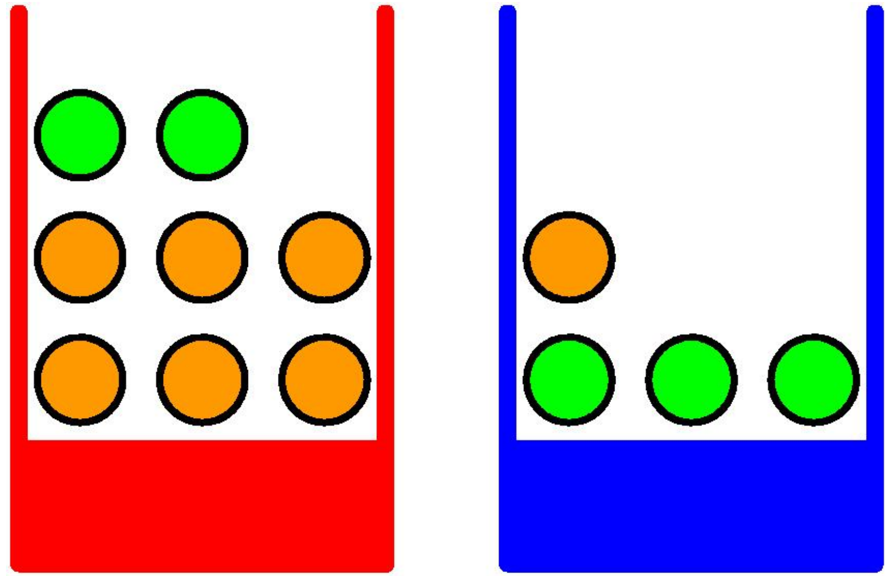
$$p(X) = \sum_Y p(X|Y)p(Y)$$

posterior ∝ likelihood × prior

Application of Prob Rules

Assume $P(Y=r) = 40\%$, $P(Y=b) = 60\%$
$P(X=a|Y=r) = 2/8 = 25\%$
$P(X=o|Y=r) = 6/8 = 75\%$

$P(X=a|Y=b) = 3/4 = 75\%$
$P(X=o|Y=b) = 1/4 = 25\%$



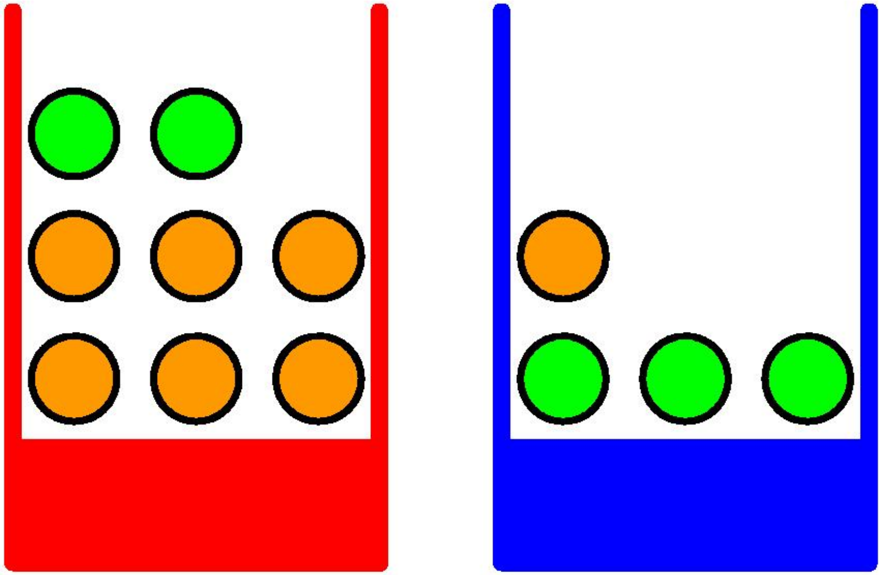
$p(X=a) = p(X=a,Y=r) + p(X=a,Y=b)$

$p(Y=r|X=o) = p(Y=r,X=o)/p(X=o)$

Application of Prob Rules

Assume $P(Y=r) = 40\%$, $P(Y=b) = 60\%$
$P(X=a|Y=r) = 2/8 = 25\%$
$P(X=o|Y=r) = 6/8 = 75\%$

$P(X=a|Y=b) = 3/4 = 75\%$
$P(X=o|Y=b) = 1/4 = 25\%$



$p(X=a) = p(X=a,Y=r) + p(X=a,Y=b)$
$= p(X=a|Y=r)p(Y=r) + p(X=a|Y=b)p(Y=b)$
$=0.25*0.4 + 0.75*0.6 = 11/20$

$P(X=o) = 9/20$

$p(Y=r|X=o) = p(Y=r,X=o)/p(X=o)$
$= p(X=o|Y=r)p(Y=r)/p(X=o)$
$= 0.75*0.4 / (9/20) = 2/3$

Mean and Variance

- The mean of a random variable X is the average value X takes.

$$\mathrm{E}[X] = \sum_x x p(x) \qquad \mathrm{E}[X] = \int x p(x) dx$$

- The variance of X is a measure of how dispersed the values that X takes are.

$$\mathrm{var}[X] = \mathrm{E}[(x - E[X])^2] = \mathrm{E}[x^2] - \mathrm{E}[x]^2$$

- The standard deviation is simply the square root of the variance.

Simple Example

- X= {1, 2} with P(X=1) = 0.8 and P(X=2) = 0.2
- Mean
  - 0.8 X 1 + 0.2 X 2 = 1.2
- Variance
  - 0.8 X (1 – 1.2) X (1 – 1.2) + 0.2 X (2 – 1.2) X (2-1.2)

Sample Mean and Variance

Given a sample set of size n, $\{x_1, \cdots, x_n\}$,

- Sample mean

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Sample variance
  - Biased

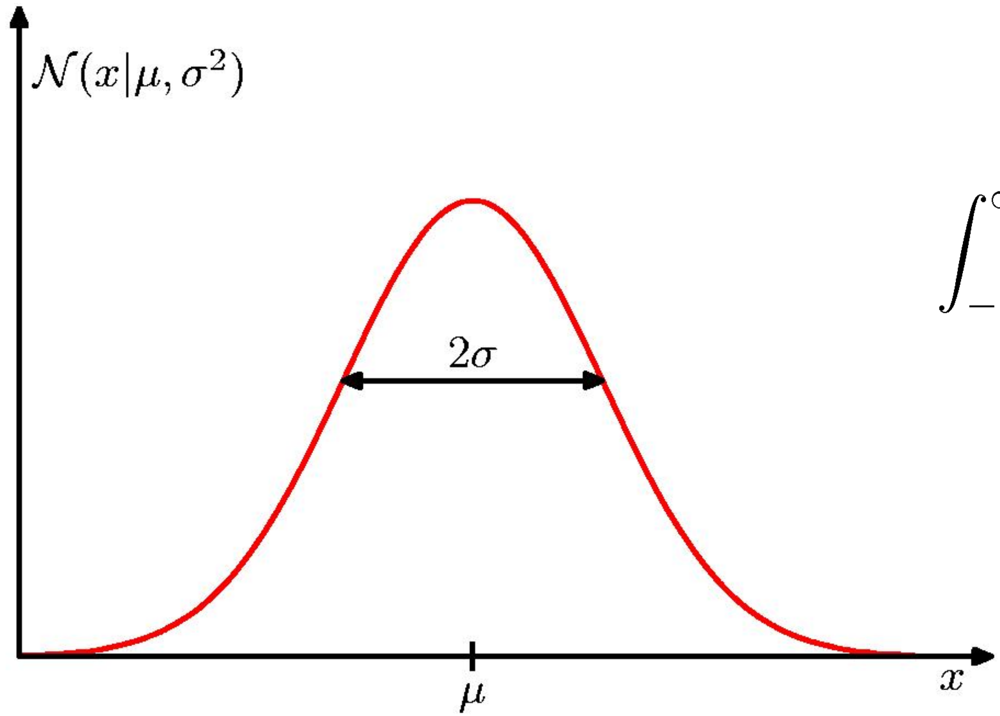
$$\sigma_y^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

  - Unbiased

$$s^2 = \frac{n}{n-1}\sigma_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

The Gaussian Distribution

$$\mathcal{N}\left(x|\mu,\sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$



$$\mathcal{N}(x|\mu,\sigma^2) > 0$$

$$\int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right) \mathrm{d}x = 1$$

Gaussian Mean and Variance

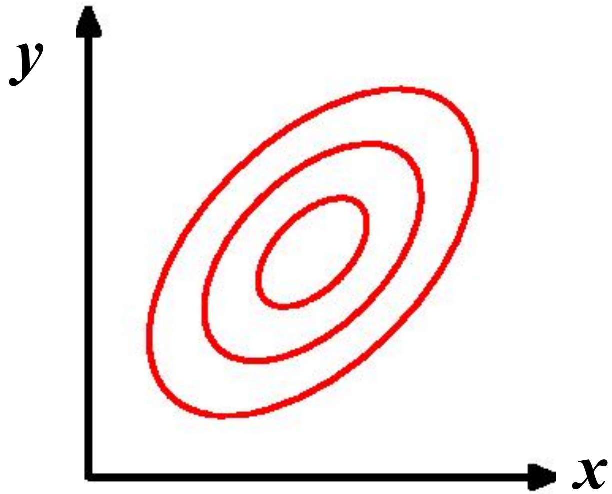
$$\mathbb{E}[x] = \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu, \sigma^2\right) x \, \mathrm{d}x = \mu$$

$$\mathbb{E}[x^2] = \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu, \sigma^2\right) x^2 \, \mathrm{d}x = \mu^2 + \sigma^2$$

$$\mathrm{var}[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = \sigma^2$$

The Multivariate Gaussian

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$



$$\begin{aligned}\operatorname{cov}[x, y] &= \mathbb{E}_{x,y}\left[\{x - \mathbb{E}[x]\}\{y - \mathbb{E}[y]\}\right] \\ &= \mathbb{E}_{x,y}[xy] - \mathbb{E}[x]\mathbb{E}[y]\end{aligned}$$

$$\begin{aligned}\operatorname{cov}[\mathbf{x}, \mathbf{y}] &= \mathbb{E}_{\mathbf{x},\mathbf{y}}\left[\{\mathbf{x} - \mathbb{E}[\mathbf{x}]\}\{\mathbf{y}^{\mathrm{T}} - \mathbb{E}[\mathbf{y}^{\mathrm{T}}]\}\right] \\ &= \mathbb{E}_{\mathbf{x},\mathbf{y}}[\mathbf{x}\mathbf{y}^{\mathrm{T}}] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}^{\mathrm{T}}]\end{aligned}$$

References

- SC_prob_basics1.pdf (necessary)
- SC_prob_basic2.pdf

  Loaded to Blackboard