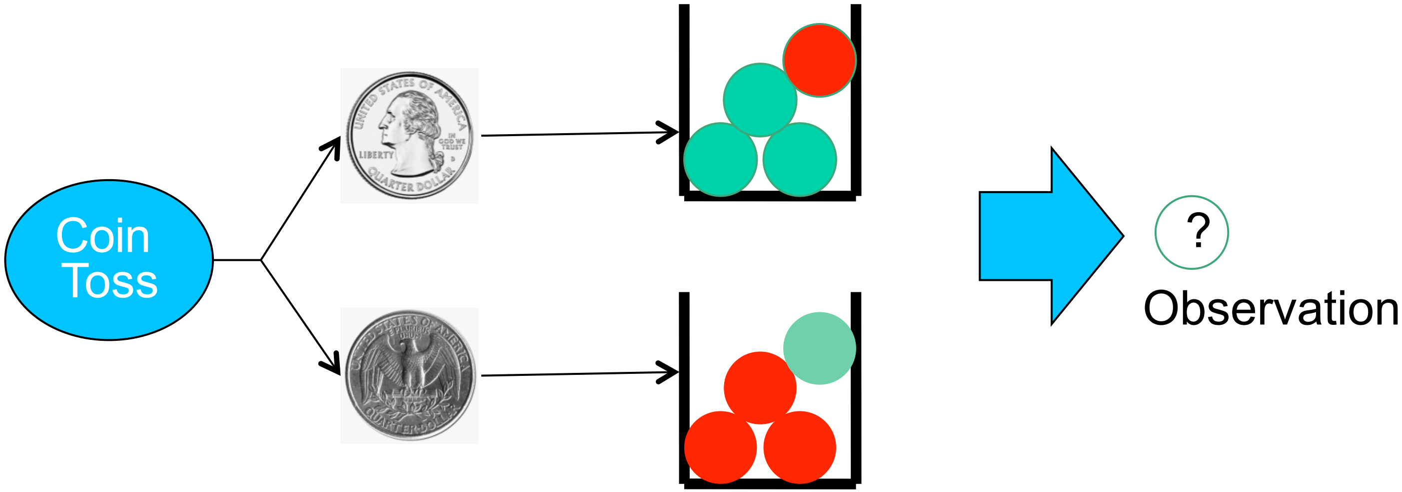
Pattern Recognition

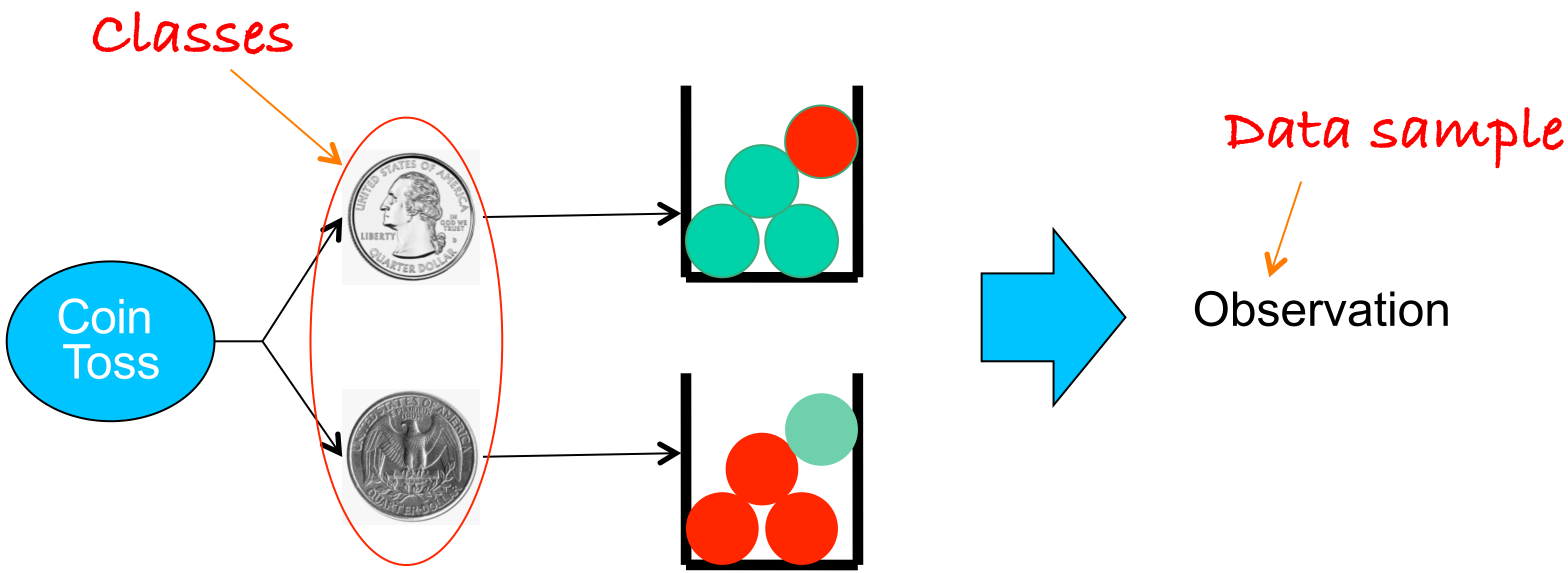


Pattern Recognition Problem

Based on the observed color of the drawn ball,
and knowing the probability of the coin toss
and composition of each of the buckets,

decide the outcome of the coin toss.

Classification Problem



Example of a Classification Problem

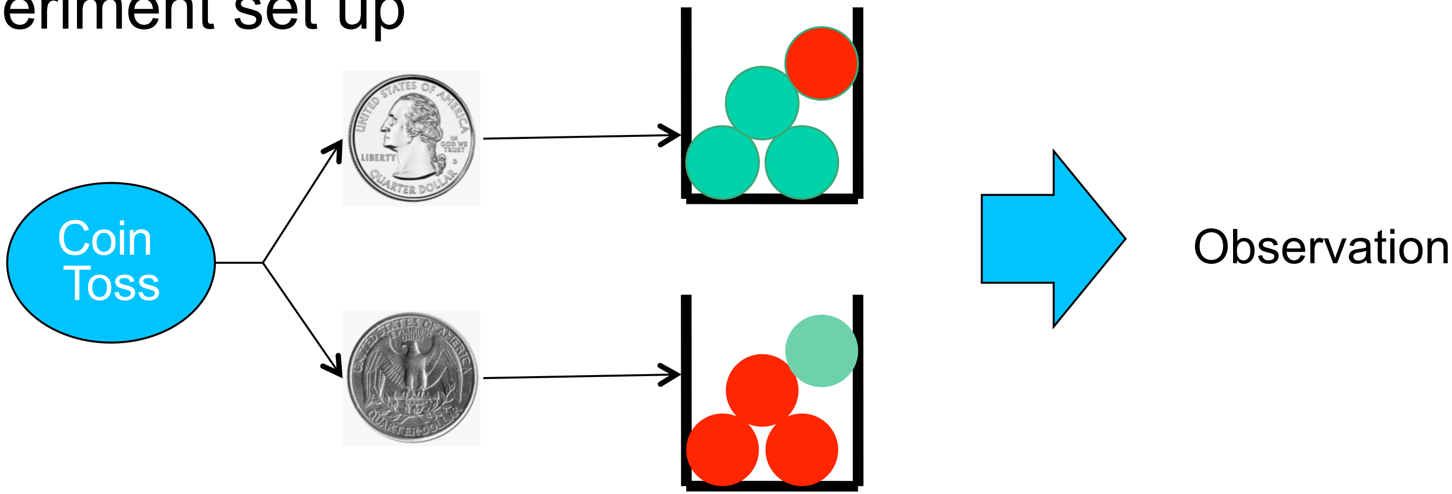
Based on
(1) the observed color of the drawn ball, and
(2) knowing the probability of the coin toss, and
(3) knowing the composition of each of the buckets,
decide the outcome of the coin toss.

Classify a data sample based on
(1) the data sample
(2) the prior probability
(3) the class densities

This is a two-class problem

Pattern Recognition Example

- Experiment set up



Decision rule

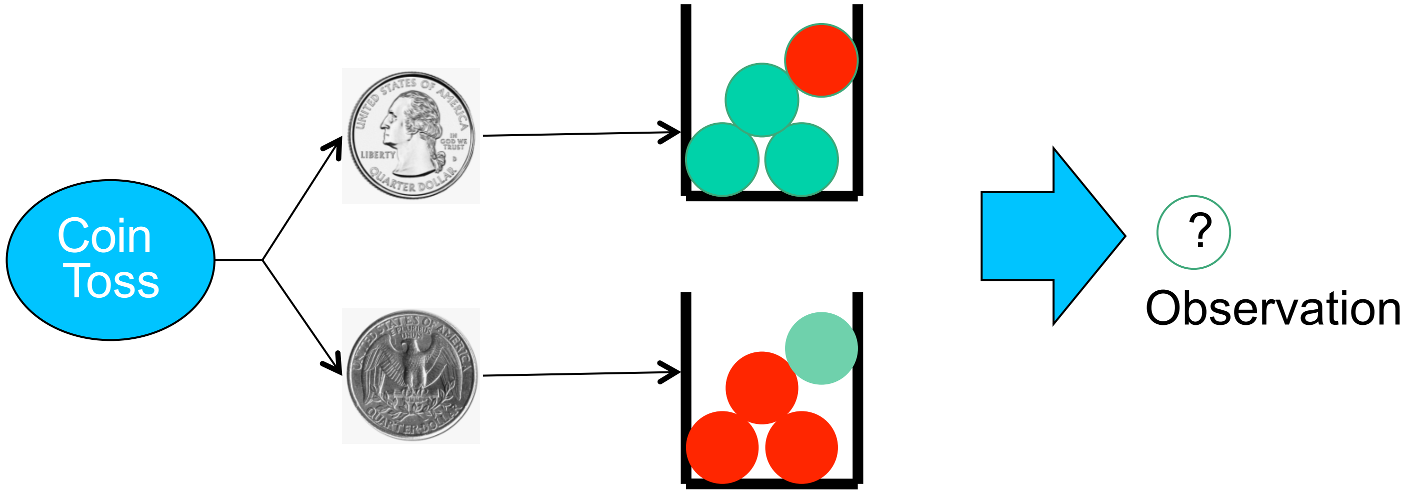
- Let us make up a decision rule

$$d(\text{observation}) = \begin{cases} \text{head} & \text{if observation is green} \\ \text{tail} & \text{if observation is red} \end{cases}$$

- Keeping score: "Suppose *A* pays *B* \$1 for every correct call, and *B* pays *A* \$1 for every incorrect call."
  - Define x as the amount of money B has
  - The expected value of x is 0.5

Payoff function

The Classification Problem



Classification Problem

Based on the observed color of the drawn ball, and knowing the probability of the coin toss and composition of each of the buckets,

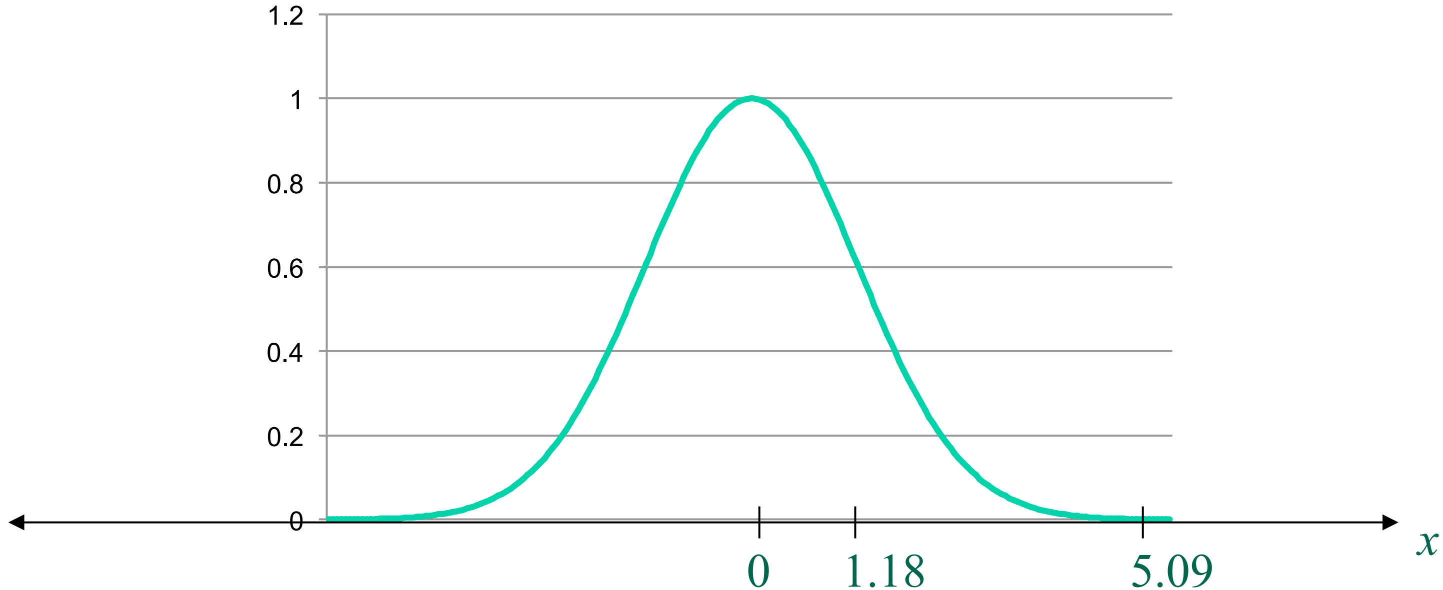
decide the outcome of the coin toss.

In practice, we do not know the exact composition of each of the buckets. Nor do we know the coin toss probability

That is why we need to do learning

Gaussian density

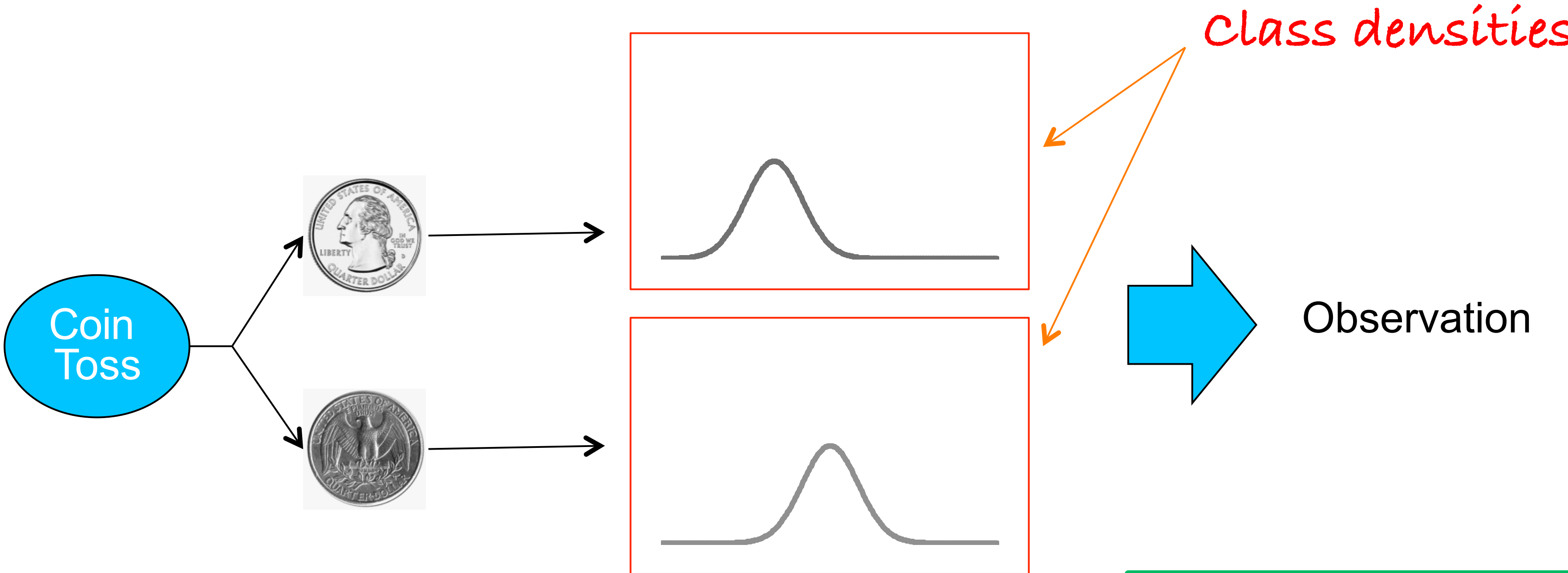
A random variable with a Gaussian distribution approximates many natural phenomena well. Its density function is a bell-shaped graph which is controlled by two basic terms: the mean and standard deviation



A Gaussian variable is a *continuous* variable. We do not talk about the probability of the variable taking on a particular value, such as $P(x = 5.09)$. Instead, we are concerned with the probability of the value in an interval, such as

$$P(\{1.18 \le x < 5.09\}) = \text{ area under the density curve from 1.18 to 5.09}$$

Another Pattern Recognition Problem



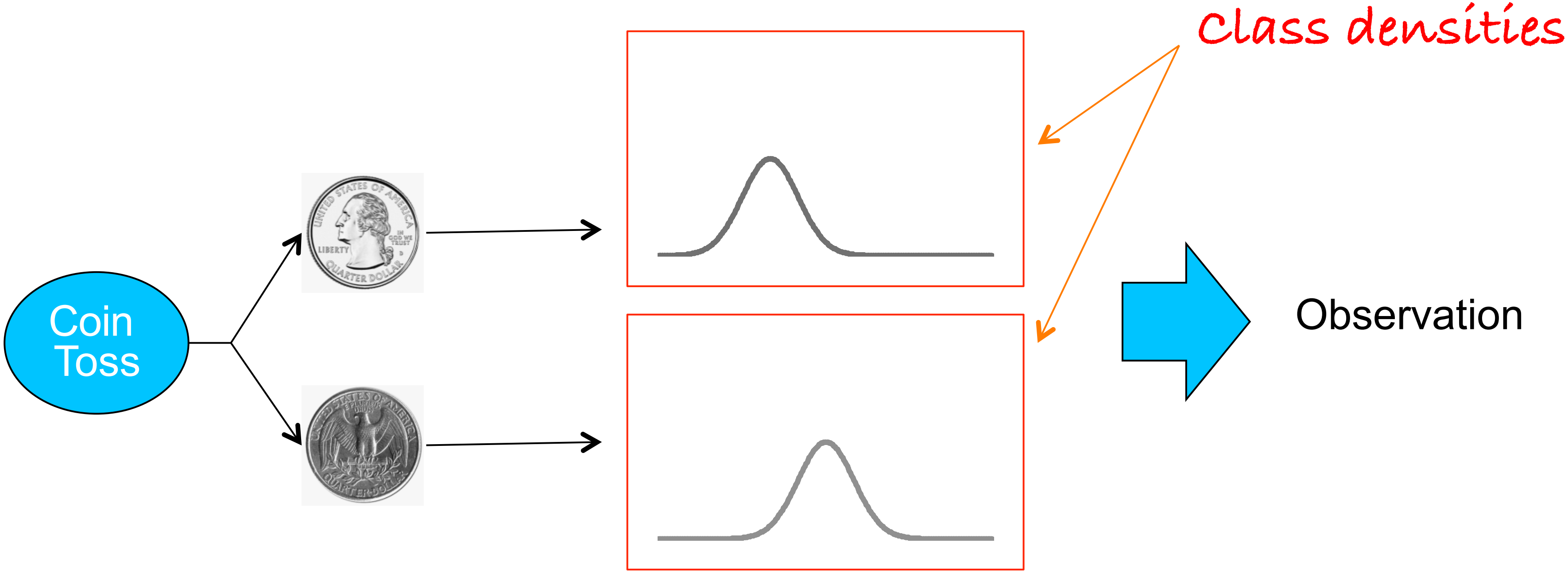
Pattern Recognition Problem

Based on the observed measurement,
and knowing the probability of the coin toss
and the densities of the measurement process,

decide the outcome of the coin toss.

In practice, we do not know the exact densities.
That is why we need to do learning

Another Pattern Recognition Problem

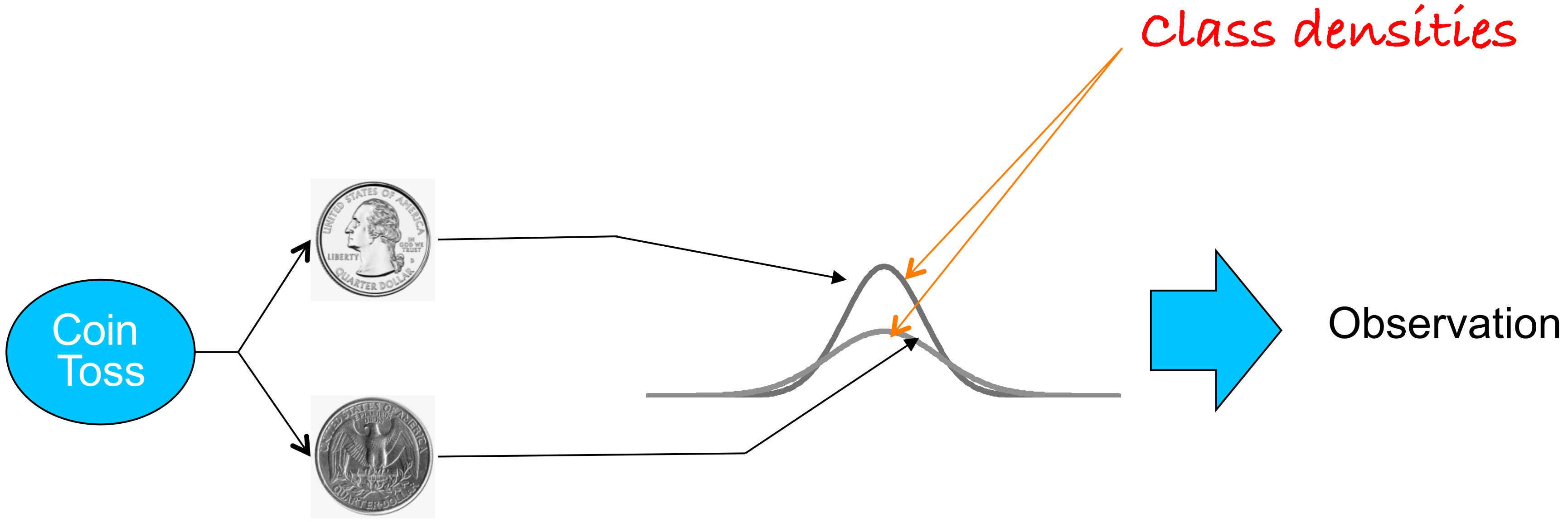


This is a common example. “Head” corresponds to the case when the class is labeled 0 and is considered “normal.” The mean is, say, 0. If “tail” the class is labeled 1 and corresponds to “abnormal.” The mean is some value non-zero, say, 1.5.

So Class 1 sample values are consistently higher than Class 0 sample values.

But of course a Class 1 sample can take the value 0. And a Class 0 sample can have a value close to 1.5

Yet Another Pattern Recognition Problem



Two class densities can have the same mean but different variances.

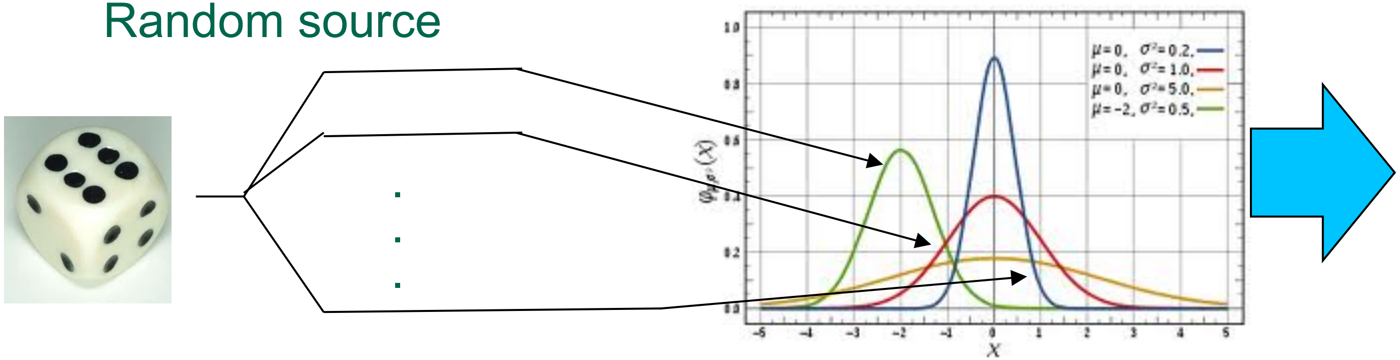
In this example, Class 1 samples have a larger variance. It is more difficult to differentiate samples from the two classes.

K-class problem

A *K*-class problem with a *K*-sided dice:

Measurement

Random source

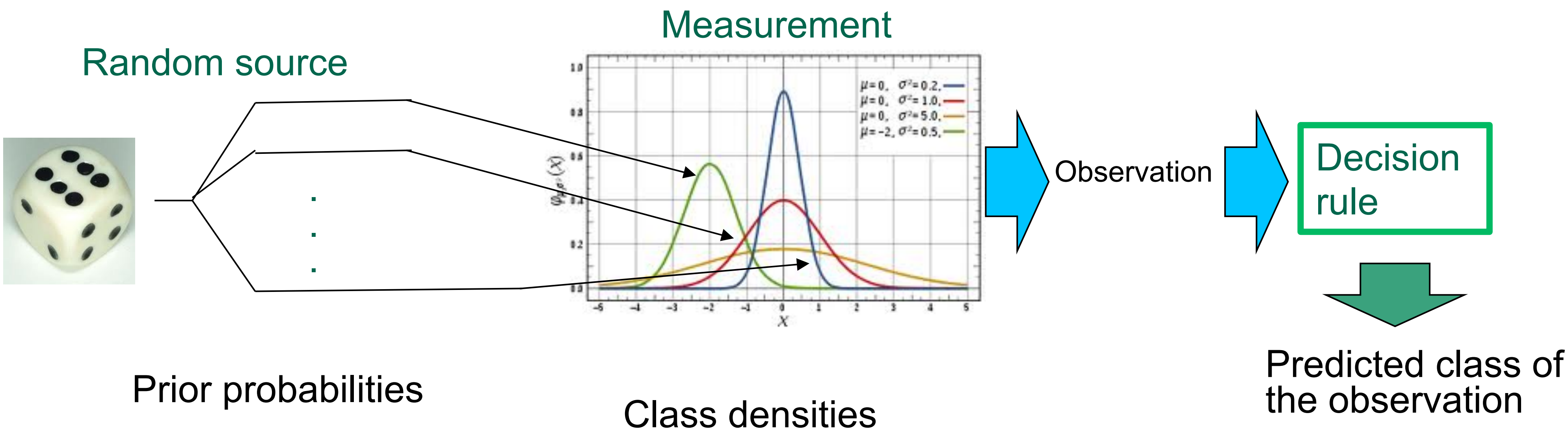


Observation

In a multi-class problem, each class has its own class density. The class densities may be from different families, differ in means or variances or both.

Optimum Decision Rule

- We want to derive an optimum decision rule
- If we know the prior probabilities and the class densities, the optimum decision rule is known



Optimal Decision Rule

- What are we optimizing?
- Candidates
  - Maximize probability of correct decision for a particular class
  - Maximize probability of correct decision for all classes

Optimal Decision Rule

Suppose *A* pays *B* \$1 for every correct call, and *B* pays *A* \$1 for every incorrect call."

Assume B is making the decision

A decision rule is optimum when the expected value of the payout is maximized

Theoretically we have to search the entire decision rule space to find the optimal one.

In practice, if the decision rule is parameterized, we optimize over the parameters

Optimal Decision Rule

Suppose *A (the "world")* penalizes *B* $1 for every incorrect call."

Assume B is making the decision

Believe it or not this is the more realistic assumption

A decision rule is optimum when the expected value of the penalty is minimized

Theoretically we have to search the entire decision rule space to find the optimal one.

In practice, if the decision rule is parameterized, we optimize over the parameters

Optimal Decision Rule

A Theory:

- Define a loss function for an incorrect decision for each class
- Define the total risk as the expected value of the total loss function (over all classes)
- The optimal decision rule that minimizes the total risk is to maximize the posterior probability

Optimal Decision Rule

A Theory:

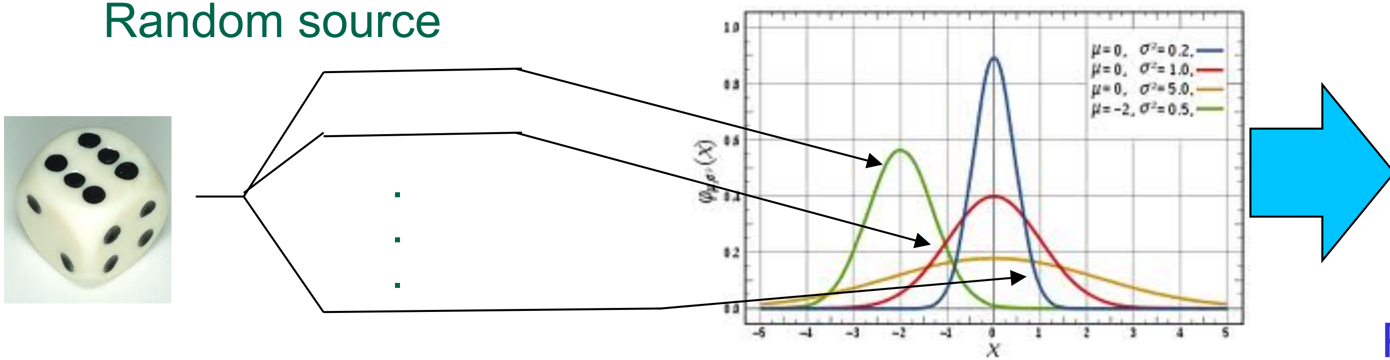
- Define a loss function for an incorrect decision for each class
- Define the total risk as the expected value of the total loss function (over all classes)
- The optimal decision rule is the one that minimizes the total risk

The optimal decision rule that minimizes the total risk is to maximize the posterior probability

Probabilities

Measurement

Random source



Observation

Posterior probability

Prior probabilities

Class k is chosen by the source with probability

$$\pi_k, k = 0, \cdots, K-1$$

Class densities

Class density k drives the observed value x

$$p(x \mid k) = p_k(x),$$

$$k = 0, \cdots, K-1$$

Given that the value x is observed, the probability that it was from Class k.

$$p(k \mid x), \; k = 0, \cdots, K-1$$

Probabilities

Prior probabilities

Class k is chosen by the source with probability

$$\pi_k, k = 0, \cdots, K-1$$

Class densities

Class density k drives the observed value x

$$p(x \mid k) = p_k(x),$$
$$k = 0, \cdots, K-1$$

Posterior probability

Given that the value x is observed, the probability that it was from Class k.

$$p(k \mid x), \; k = 0, \cdots, K-1$$

$$\underbrace{\hspace{10em}}$$

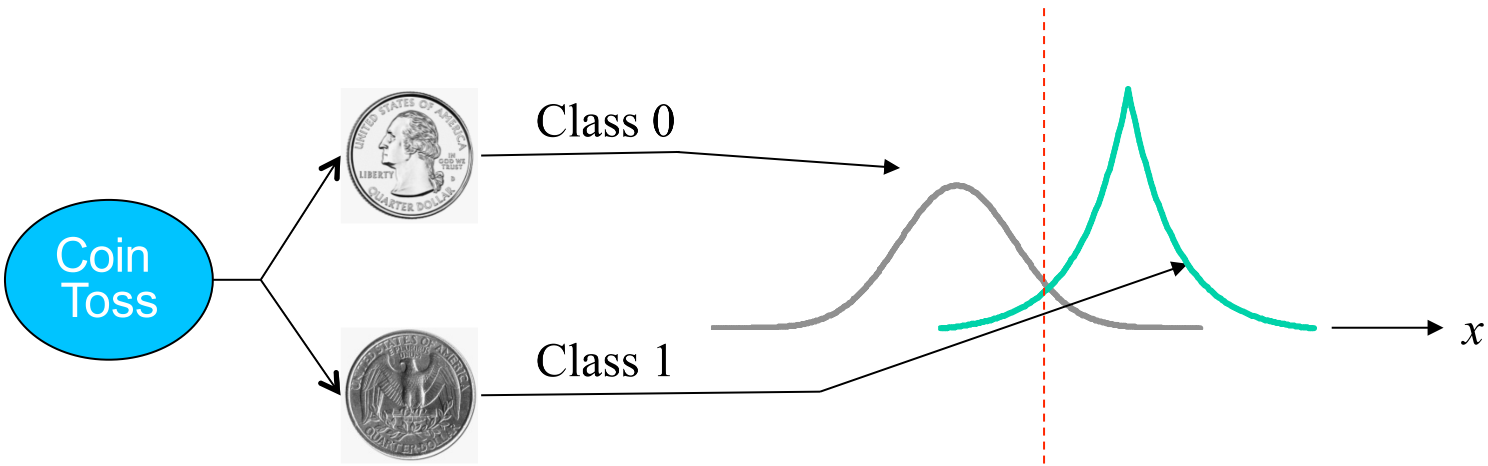
$$p(k \mid x) = \frac{p(k, x)}{p(x)} = \frac{p(x \mid k)\pi_k}{p(x)}$$

Maximum a posteriori rule says:

Choose Class k if $p(k \mid x) > p(k' \mid x)$ for all $k' \neq k$;

i.e., choose Class k if $p(x \mid k)\pi_k > p(x \mid k')\pi_{k'}$ for all $k' \neq k$;

Optimal Decision Rule



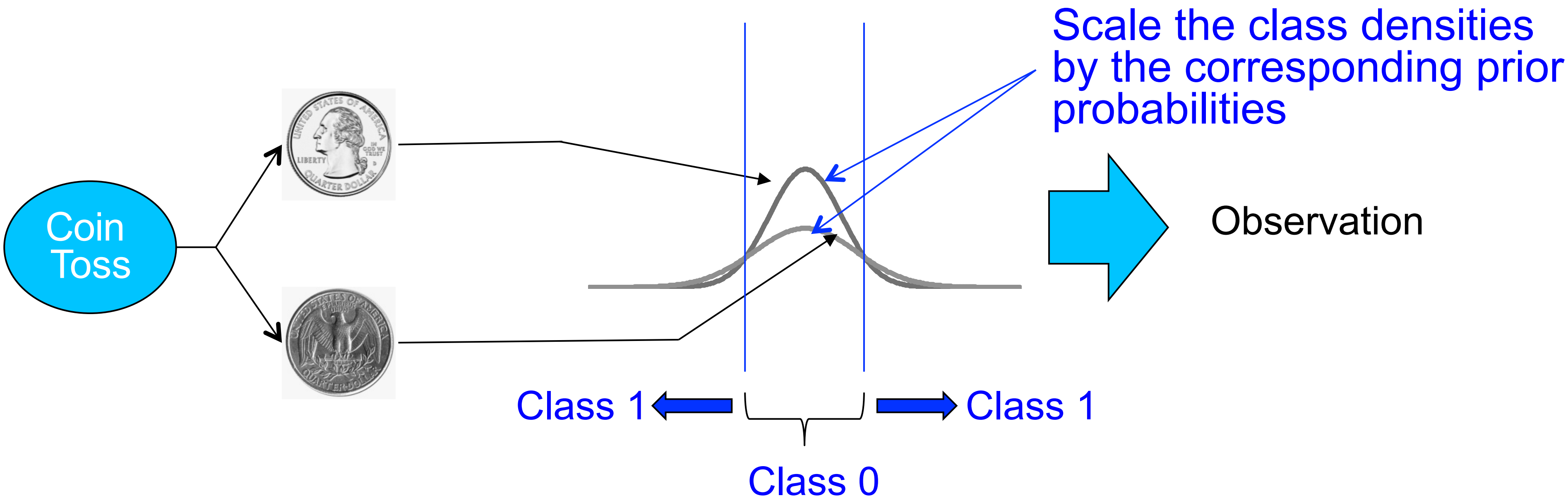
Bayesian Rule

Decide Class 0 if x is on this side | Decide Class 1 if x is on this side

Optimal Decision Rule

Maximum a posteriori rule says:

Choose Class k if $p(x \mid k)\pi_k > p(x \mid k')\pi_{k'}$ for all $k' \neq k$;



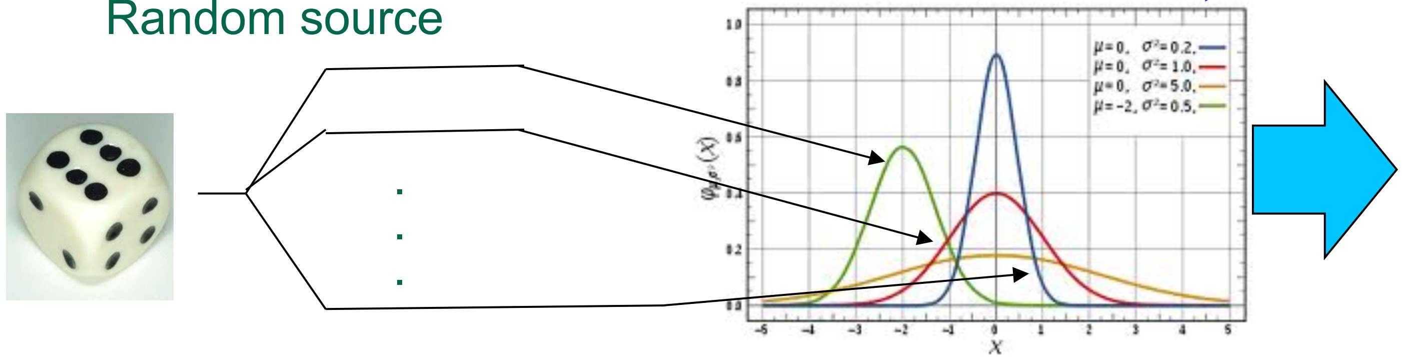
K-class problem

A K-class problem with a K-sided dice:

Scale the class densities by the corresponding prior probabilities

Measurement

Random source



Observation

Maximum a posteriori rule says:

Choose Class k if $p(x \mid k)\pi_k > p(x \mid k')\pi_{k'}$ for all $k' \neq k$;

Optimal Decision Rule

- Identify in the classification problem
  - Number of classes
  - Prior probabilities
  - Class densities
  - The observation sample variable
- Optimal decision rule
  - Is defined for the whole range of the observation sample variable
  - Is given by the Maximum a posteriori rule

$$\text{Choose Class } k \text{ if } p(x \mid k)\pi_k > p(x \mid k')\pi_{k'} \text{ for all } k' \neq k;$$