Bayesian Decision Theory

- **fundamental statistical approach to pattern classification using**
  – probability of classification
  – cost of error

Sample classification scenario

- **The CMU Robotics Institute has built an autonomous robot for NASA to search for meteorites in Antarctica**
  – there are lots of meteorites in Antarctica; they fall, land on the ice, and stay
  – the environment is too inhospitable for human researchers to retrieve them
  – practice for Moon, Mars

Sample classification scenario

- **The robot's rock detector goes off**
- **There's either a terrestrial rock or a meteorite**
- **Questions now:**
  - What does the robot conclude?
  - What should the robot do?

Formal description

- **Nature is in one of two states**
  - variable ω: state of nature
  - value $\omega=\omega_1$: earth rock
  - value $\omega=\omega_2$: space rock

Formal description
CTD

- **As state of nature is so unpredictable, we describe variable ω probabilistically**
  - A priori probabilities (priors)

$$P(\omega_1)$$
$$P(\omega_2)$$

  - Positive, sum to one
  - Specify our knowledge of how likely any Antarctic rock is to be from earth or space

Decision Rules

- **Say the robot must decide on the rock (without knowing anything else about it)**
- **Probabilistic decision rule**

$$\text{decide } \omega_1 \text{ if } P(\omega_1) > P(\omega_2)$$
$$\text{otherwise decide } \omega_2$$

Risk

- **NASA didn't send you all the way to Antarctica to sit on the tundra and sulk**
- **Two possible actions**
  - α_1: leave rock alone
  - α_2: pick it up
- **Loss associated with action in state**

  $\lambda(\alpha_i \mid \omega_j)$ - abbrev: λ_{ij}

  - For now, assume $\lambda_{ii} = 0$

Risk
CTD

- **Risk is expected loss, here**

  $$R(\alpha_i) = \lambda_{i1}\, P(\omega_1) + \lambda_{i2}\, P(\omega_2)$$

- **Choose action to minimize risk**

  If $R(\alpha_1) < R(\alpha_2)$ then do α_1;
  otherwise do α_2

- **Concretely:**

  If $\lambda_{12}\, P(\omega_2) < \lambda_{21}\, P(\omega_1)$ then do α_1;
  otherwise do α_2

Adding some evidence

- **First case: continuous measurement**
- **Example, for Antarctic robot**
  - – Visual rock detector gives you back an estimate of the redness of the rock
  - – Meteors tend to be redder than earth rocks (because they're more likely ferrous)
  - – So redness is useful information

Formalism

- **Measurement *x***
- **Class-conditional probability density fn**

$$p(x \mid \omega_j)$$

  - – assumes nature is in ω_j
  - – describes for each possible measurement x its likelihood relative to other possible measurements

$$\int p(x \mid \omega_j)\, dx = 1$$

Problem statement

- **Suppose we know**
  - Priors $P(\omega_j)$ (for each j)
  - Likelihood $p(x \mid \omega_j)$ (for each j)
  - Measurement x
- **How does this influence our attitude concerning the true state of nature?**

Answer
PART 1

- **Whatever ω is, say ω_j, it's combined with x now - as characterized by density**

$$p(\omega_j, x)$$

- **We can understand this in two ways**
  - from x, determine ω_j

$$p(\omega_j, x) = P(\omega_j \mid x)\ p(x)$$

  - from ω_j, determine x

$$p(\omega_j, x) = p(x \mid \omega_j)\ P(\omega_j)$$

Answer
PART 2

- **We only know we have *x*; we want to compare alternative**
- **From before**

$$P(\omega_j \mid x)\ p(x) = p(x \mid \omega_j)\ P(\omega_j)$$

- **Thus**

$$P(\omega_j \mid x) = \frac{p(x \mid \omega_j) P(\omega_j)}{p(x)}$$

- **Bayes's formula**

posterior = likelihood × prior ⁄ evidence

Bayes Decision Rule

- **Algorithm for minimizing expected error**
  - in binary statistical decision
- **Given measurement *x***
- **If $P(\omega_1 \mid x) > P(\omega_2 \mid x)$**
  - decide ω_1
- **Otherwise**
  - decide ω_2

Justification

- **In any case**

$$P(\text{error} \mid x) = \begin{cases} P(\omega_1 \mid x) \text{ if we decide } \omega_2 \\ P(\omega_2 \mid x) \text{ if we decide } \omega_1 \end{cases}$$

- **Overall**

$$P(\text{error}) = \int P(\text{error},x)\ dx$$
$$= \int P(\text{error} \mid x)\ p(x)\ dx$$

- **Our algorithm makes P(error | x) as small as possible, which minimizes integral here**

A Step Back

- **By Bayes's formula, decision is**

$$\frac{p(x \mid \omega_1)P(\omega_1)}{p(x)} > \frac{p(x \mid \omega_2)P(\omega_2)}{p(x)}$$

- **Scale factor $p(x)$ has no impact on decision:**

$$p(x \mid \omega_1)\ P(\omega_1) > p(x \mid \omega_2)\ P(\omega_2)$$

A Step Back

- **Two cases for:**

$$p(x \mid \omega_1)\, P(\omega_1) > p(x \mid \omega_2)\, P(\omega_2)$$

- **No info from test:**

$$\text{decide } P(\omega_1) > P(\omega_2)$$

$$p(x \mid \omega_1) = p(x \mid \omega_2)$$

- **No background preference:**

$$\text{decide } p(x \mid \omega_1) > p(x \mid \omega_2)$$

$$P(\omega_1) = P(\omega_2)$$